Convergence and Limit Theorems

Paweł Polak

January 26, 2016

STAT W4413: Nonparametric Statistics - Lecture 4

Convergence and limit theorems

- Calculating the significance level and probability of Type II error of most of the tests that we see in this lecture, requires nontrivial calculations.

- One of the main tools that will help us in such cases is the asymptotic performance analysis.

- With very little computational power, such analysis can usually provide sufficiently accurate results, especially when the number of samples is high.

Convergence in distribution

Definition

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables with $X_n \sim F_n$. Also, let $X \sim F$. The sequence $\{X_n\}_{n=1}^{\infty}$ is said to converge to X in distribution, denoted by

$$X_n \xrightarrow{d} X,$$

if and only if

$$F_n(x) \to F(x)$$

for all continuity points of F as $n \to 1$.

Note that it is usually easier to prove that the probability density function of X_i converges to the probability density function of X.

Remark

It turns out that if the pdf of X_n converges to the pdf of X, then we can conclude that $X_n \xrightarrow{d} X$.

- This is a simple corollary of Scheffe's lemma.
- The same is true for the probability mass functions, i.e., if you prove the convergence of the probability mass functions, then you again have the convergence in distribution.

Example

Example

Let $X_i \sim N(\frac{1}{n}, 1)$. What is the limiting distribution of $\{X_i\}_{i=1}^{\infty}$? It is straightforward to confirm that the pdf of X_i is converging to the pdf of $N(0, 1)$. Therefore, $N(0, 1)$ is the limiting distribution.

Example

A well-known example of convergence in distribution is the *Central Limit Theorem (CLT)*.

Theorem

Let $X_1, X_2, \ldots \overset{i.i.d.}{\sim} F$*, with* $\mathbb{E}(X_i) = \mu < \infty$ *and* $Var(X_i) = \sigma^2 < \infty$*.* *Define the sample average as* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$*. Then*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Example

Example

Let X_n be a $\chi^2(n)$. What is the limiting distribution of $Y_n = \frac{1}{\sqrt{n}}(X_n - n)$?

Example

Let $Z_1, Z_2, \ldots, Z_n, \ldots \overset{iid}{\sim} N(0,1)$. Since X_n is $\chi^2(n)$, its distribution is the same as the distribution of $\sum_{i=1}^n Z_i^2$, which is the sum of i.i.d. random variables Z_i^2 . Note that $\mathbb{E}(Z_i^2) = 1$. Also,

$$var(Z_i^2) = \mathbb{E}(Z_i^2 - 1)^2 \overset{(a)}{=} 2.$$

To obtain Equality (a), we have used the fact that $\mathbb{E}(Z_i^4) = 3$. Can you prove it? Now it is clear from CLT that the random variable $\sqrt{n}(\frac{1}{n}(X_n) - 1) \to N(0,2)$. It is also straightforward to confirm that if $Y_n = \frac{1}{\sqrt{n}}(X_n - n)$, then $Y_n \xrightarrow{d} N(0,2)$.

Convergence in probability

Definition

Consider a sequence of random variables $\{X_i\}_{i=1}^{\infty}$. The sequence is said to converge to X in probability, denoted by $X_n \xrightarrow{p} X$, if and only if $\forall \epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1.$$

Convergence in probability

Example

Consider $X_n \sim \text{Exponential}(\lambda n)$, i.e., the pdf of X_n is

$$f_{X_n}(x) = \lambda n e^{-\lambda n x} \mathbb{I}(x \geq 0),$$

and $\lambda > 0$ is fixed and $\mathbb{I}$ is the indicator function.
Prove that $X_n \xrightarrow{p} 0$ as $n \to \infty$.

Proof.

In order to prove this convergence we should calculate $\mathbb{P}(|X_n| > \epsilon)$. It is straightforward to confirm that

$$\mathbb{P}(|X_n| > \epsilon) = e^{-\lambda n \epsilon} \underset{n\to\infty}{\to} 0.$$

Hence $X_n \xrightarrow{p} 0$. $\square$

Convergence in probability: WLLN

A well-known example of convergence in probability, that will be used in this course often, is the *weak law of large numbers (WLLN)*.

Theorem (WLLN)

Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. with $E(X_i) = \mu < \infty$ and $var(X_i) = \sigma^2 < \infty$. Then $\overline{X}_n \xrightarrow{p} \mu$

- The proof of this statement is straightforward and I encourage you to prove this.
- The only inequality that you may need to prove this result is Markov inequality (or more specifically, Chebyshev.
- You can look it up in Wikipedia to refresh your memories.

Almost sure convergence

The last type of convergence we discuss here is the *almost sure convergence*.

Definition

$\{X_n\}_{n=1}^{\infty}$ *is said to converge to* X *almost surely denoted by* $X_n \xrightarrow{a.s.} X$ *if and only if*

$$\mathbb{P}(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}) = 1.$$

Almost sure convergence is also known as *convergence with probability one*.

Almost sure convergence: SLLN

A well-known example of almost sure convergence is the *Strong Law of Large Numbers*.

Theorem (SLLN)

Let $X_1, X_2, \ldots, X_n, \ldots$ *be iid with* $\mathbb{E}(X_i) = \mu < \infty$ *and* *var*$(X_i) = \sigma^2 < \infty$. *Then* $\overline{X}_n \stackrel{a.s.}{\to} \mu$.

Note that among the three types of convergence we have discussed so far, almost sure convergence is the strongest. In fact if $X_n \stackrel{a.s.}{\to} X$ then X_n converges to X in both probability and distribution. We discuss some parts of this statement in the next slides and will leave the rest as an exercise for you.

Connection of convergence in probability and convergence in distribution

Theorem

$$\textit{If } X_n \xrightarrow{p} X, \ \textit{then } X_n \xrightarrow{d} X.$$

Connection of convergence in probability and convergence in distribution

Proof.

We would like to prove that $\lim_{n\to\infty} \mathbb{P}(X_n \le a) = \mathbb{P}(X \le a)$. Let $\epsilon > 0$ be a fixed number. Here we employ a very useful trick known as conditioning. We have

$$\mathbb{P}(X_n \le a) = \mathbb{P}(X_n \le a \mid |X_n - X| < \epsilon)\mathbb{P}(|X_n - X| < \epsilon) + \mathbb{P}(X_n \le a \mid |X_n - X| \ge \epsilon)\mathbb{P}(|X_n - X| \ge \epsilon) \tag{1}$$

We now use (1) to find an upper bound and a lower bound for $\mathbb{P}(X_n \le a)$. Let's start with the upper bound.

$$\begin{aligned}
\mathbb{P}(X_n \le a) &\overset{(b)}{\le} \mathbb{P}(X_n \le a \mid |X_n - X| < \epsilon)\mathbb{P}(|X_n - X| < \epsilon) + \mathbb{P}(|X_n - X| \ge \epsilon) \\
&\overset{(c1)}{\le} \mathbb{P}(X \le a + \epsilon \mid |X_n - X| < \epsilon)\mathbb{P}(|X_n - X| < \epsilon) + \mathbb{P}(|X_n - X| \ge \epsilon) \\
&\overset{(c2)}{\le} \mathbb{P}(X \le a + \epsilon,\ |X_n - X| < \epsilon) + \mathbb{P}(|X_n - X| \ge \epsilon) \\
&\overset{(d)}{\le} \mathbb{P}(X \le a + \epsilon) + \mathbb{P}(|X_n - X| \ge \epsilon) \qquad (2)
\end{aligned}$$

□

Connection of convergence in probability and convergence in distribution

Proof.

To obtain a lower bound we again use (1). Since $\mathbb{P}(X_n \leq a|\ |X_n - X| \geq \epsilon)\mathbb{P}(|X_n - X| \geq \epsilon) \geq 0$, we have

$$\begin{aligned}\mathbb{P}(X_n \leq a) &\geq \mathbb{P}(X_n \leq a \mid\ |X_n - X| < \epsilon)\mathbb{P}(|X_n - X| < \epsilon) \\ &\overset{(e)}{\geq} \mathbb{P}(X \leq a - \epsilon \mid\ |X_n - X| < \epsilon)\mathbb{P}(|X_n - X| < \epsilon) \\ &\overset{(f)}{=} \mathbb{P}(X \leq a - \epsilon) - \mathbb{P}(X \leq a - \epsilon|\ |X_n - X| \geq \epsilon)\mathbb{P}(|X_n - X| \geq \epsilon) \\ &\overset{(g)}{\geq} \mathbb{P}(X \leq a - \epsilon) - \mathbb{P}(|X_n - X| \geq \epsilon) \qquad (3)\end{aligned}$$

We obtain Inequality (e) by the following argument: $|X_n - X| < \epsilon$ implies that $X_n \leq X + \epsilon$. Also if we replace X_n with something larger $\mathbb{P}(X_n \leq a|\ |X_n - X| < \epsilon)$ decreases. Hence we replace X_n with $X + \epsilon$ and obtain Inequality (e). Equality (f) is due to the fact that

$$\mathbb{P}(X \leq a-\epsilon) = \mathbb{P}(X \leq a-\epsilon|\ |X_n-X| \geq \epsilon)\mathbb{P}(|X_n-X| \geq \epsilon)+\mathbb{P}(X \leq a-\epsilon|\ |X_n-X| < \epsilon)\mathbb{P}(|X_n-X| < \epsilon).$$

Finally, to obtain Inequality (g) from $\mathbb{P}(X \leq a - \epsilon|\ |X_n - X| \geq \epsilon) \leq 1$. □

Connection of convergence in probability and convergence in distribution

Proof.

Combining (2) and (3) we obtain

$$\mathbb{P}(X \leq a - \epsilon) - \mathbb{P}(|X_n - X| \geq \epsilon) \leq \mathbb{P}(X_n \leq a) \leq \mathbb{P}(X \leq a + \epsilon) + \mathbb{P}(|X_n - X| \geq \epsilon).$$

Letting $n \to \infty$ implies that (since $X_n \xrightarrow{p} X$) for every $\epsilon \geq 0$.

$$\mathbb{P}(X \leq a - \epsilon) \leq \lim_{n \to \infty} \mathbb{P}(X_n \leq a) \leq \mathbb{P}(X \leq a + \epsilon). \tag{4}$$

If a is a continuity point of F, then letting $\epsilon \to 0$ establishes the result (by using Sandwich theorem for limits). $\square$

Remarks

- While convergence in probability implies convergence in distribution, the other direction is not correct. i.e., convergence in distribution does not imply convergence in probability. Here is an example:

Example

Let $X \sim Unif(0,1)$. We define $\{X_n\}_{n=1}^{\infty}$, with $X_n = X$. First note that $X_n \xrightarrow{d} X$. Why?

Since X has the same distribution as $1 - X$, we can conclude that $X_n \xrightarrow{d} 1 - X$. However, I claim X_n does not converge to $1 - X$ in probability. Because,

$$\mathbb{P}(|X_n - 1 + X| \geq \epsilon) = \mathbb{P}(|X - 1 + X| \geq \epsilon) = 1 - \epsilon.$$

- Make sure to calculate the probability yourself and prove that $\mathbb{P}(|X - 1 + X| \geq \epsilon) = 1 - \epsilon$.
- As this example shows the convergence in distribution does not imply the convergence in probability. That said, there is a very special but important partial inverse for the last Theorem that we will mention in the next Theorem.

Remarks

Theorem

Let $c \in \mathbb{R}$ be a fixed (nonrandom) number. If $X_n \xrightarrow{d} c$, then $X_n \xrightarrow{p} c$.

Proof.

Let $F(x)$ be defined in $F(x) \triangleq \mathbb{I}(x \geq c)$.

$$\mathbb{P}(|X_n - c| \leq \epsilon) = \mathbb{P}(c - \epsilon \leq X_n \leq c + \epsilon) \geq \mathbb{P}(c - \epsilon < X_n \leq c + \epsilon) = F_{X_n}(c + \epsilon) - F_{X_n}(c - \epsilon) \quad (5)$$

Clearly F_{X_n} converges to F and the only discontinuity point of F occurs at $x = c$. Therefore,

$$\begin{aligned} \lim_{n \to \infty} F_{X_n}(c + \epsilon) &= F(c + \epsilon) = 1 \\ \lim_{n \to \infty} F_{X_n}(c - \epsilon) &= F(c - \epsilon) = 0 \end{aligned} \quad (6)$$

Combining (5) with (6), we obtain

$$\lim_{n \to \infty} \mathbb{P}(|X_n - c| \leq \epsilon) \geq 1.$$

But we also know that $\mathbb{P}(|X_n - c| \leq \epsilon) \leq 1$. Therefore, $\lim_{n \to \infty} \mathbb{P}(|X_n - c| \leq \epsilon) = 1$, that completes the proof. □

Some useful results

To obtain the limiting distributions of different estimators and different tests, we usually need to combine different convergence results. Here is an example to clarify this statement:

Example

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$, with mean μ and variance σ^2, where both μ and σ^2 are unknown. We estimate the variance by

$$\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}^2, \tag{7}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. (Can you prove the last equality?)
The question is what does $\bar{\sigma}^2$ converge to in probability. If we look at the right hand side of (7) we notice that the term $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ converges in probability to $\mathbb{E}(X_i^2)$. Why?
Furthermore, $\bar{X} \to \mathbb{E}(X)$ in probability.

The question is can we employ these individual results to obtain a new result regarding $\bar{\sigma}^2$? In other words, can we say that $\bar{\sigma}^2 \xrightarrow{p} \mathbb{E}(X^2) - (\mathbb{E}(X))^2$? Furthermore, can we employ CLT for instance, to characterize the distribution of $\sqrt{n}(\bar{\sigma}^2 - \sigma^2)$? In the next few slides, we introduce some tools that enable you to do these calculations.

Some useful results

- Suppose that we have two sets of random variables

$$X_n \xrightarrow{d} X \text{ and } Y_n \xrightarrow{d} Y.$$

- Can we characterize the limiting distribution of $X_n + Y_n$?

Some useful results

- The answer is NO.
- This is due to the fact that marginal distributions of X_n and Y_n does not include all the information about their joint behavior.

Example

For instance, suppose that $X_n \xrightarrow{d} N(0,1)$. Then define $Y_n = -X_n$. It is clear that $Y_n \xrightarrow{d} N(0,1)$ as well (Can you prove this claim?). However, $X_n + Y_n \xrightarrow{d} 0$. Now if we define a new set of random variables $W_n = X_n$, then $X_n + W_n \xrightarrow{d} 2N(0,1)$. Why? Therefore, as long as we do not have a good understanding of the joint distribution of (X_n, Y_n), we cannot derive the limiting distribution of $X_n + Y_n$.

However,...

Slutsky's theorem

Slutsky's theorem is a special case for which we can characterize the distribution of $X_n + Y_n$ from the limiting distribution of X_n and the limiting distribution of Y_n only.

Theorem (Slutsky's theorem)

Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, where X is a random variable, but $c \in \mathbb{R}$ is a fixed number. Then,

1. $X_n + Y_n \xrightarrow{d} X + c$,
2. $X_n Y_n \xrightarrow{d} cX$,
3. *If $c \neq 0$, then* $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$.

Slutsky's theorem

We only prove part (1) here. (You will prove the rest in the next Homework.)

Proof.

First note that $Y_n \xrightarrow{d} c$ implies $Y_n \xrightarrow{p} c$. Keeping that in mind, we use the conditioning argument we discussed before and we provide an upper bound and a lower bound for $\mathbb{P}(X_n + Y_n \leq t)$.

$$\begin{aligned}
\mathbb{P}(X_n + Y_n \leq t) &\overset{(a)}{=} \mathbb{P}(X_n + Y_n \leq t \mid |Y_n - c| \leq \epsilon)\mathbb{P}(|Y_n - c| \leq \epsilon) + \mathbb{P}(X_n + Y_n \leq t \mid |Y_n - c| > \epsilon)\mathbb{P}(|Y_n - c| > \epsilon) \\
&\overset{(b)}{\leq} \mathbb{P}(X_n + Y_n \leq t \mid |Y_n - c| \leq \epsilon)\mathbb{P}(|Y_n - c| \leq \epsilon) + \mathbb{P}(|Y_n - c| > \epsilon) \\
&\overset{(c)}{\leq} \mathbb{P}(X_n \leq -c + \epsilon + t \mid |Y_n - c| \leq \epsilon)\mathbb{P}(|Y_n - c| \leq \epsilon) + \mathbb{P}(|Y_n - c| > \epsilon) \\
&\overset{(d)}{\leq} \mathbb{P}(X_n \leq -c + \epsilon + t) + \mathbb{P}(|Y_n - c| > \epsilon)
\end{aligned} \tag{8}$$

You will argue in Homework, why each of the equalities or inequalities above are right. Furthermore, you will prove in the next Homework that

$$\mathbb{P}(X_n + Y_n \leq t) \geq \mathbb{P}(X_n \leq -c - \epsilon + t) - \mathbb{P}(|Y_n - c| \geq \epsilon). \tag{9}$$

If we combine (8) and (9) and let $n \to \infty$, we obtain

$$\mathbb{P}(X \leq -c - \epsilon + t) \leq \lim_{n\to\infty} \mathbb{P}(X_n + Y_n \leq t) \leq \mathbb{P}(X \leq -c + \epsilon + t) \; \forall \epsilon > 0. \tag{10}$$

If we let $\epsilon \to 0$ we obtain the result. □

$P + P \Rightarrow P$

As we discussed before Slutsky's theorem, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ does not imply anything about the distribution of $X_n + Y_n$. The next theorem says that $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ in fact imply that $X_n + Y_n \xrightarrow{p} X + Y$.

Theorem

If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then

$$X_n + Y_n \xrightarrow{p} X + Y.$$

- You will prove this in HW.
- Note that this result can be extended to the products of two random variables and if $P(Y = 0) = 0$, then we can say $X_n/Y_n \xrightarrow{p} X/Y$.

One of the issues that led us to Slustky's theorem was that we could not say anything about the limiting distribution of $X_n + Y_n$ even though we knew the limiting distribution of X_n and Y_n individually. To be able to say things about the distribution $X_n + Y_n$ we have to study the joint distribution of these two random variables. This leads us to the study of *convergence of random vectors*.

Convergence in distribution for random vectors

Definition

Let $X_1, X_2, \ldots, X_n, \ldots$ be random vectors in $\mathbb{R}^d$. X_n converges to X in distribution, denoted by $X_n \xrightarrow{d} X$, if and only if the joint distribution of $F_{X_n}(a) \to F_X(a)$ for any $a \in \mathbb{R}^d$ that is a continuity point of $F_X(a)$.

An important instance of convergence in distribution is the central limit theorem (CLT).

Theorem

Let $X_1, X_2, \ldots, X_n, \ldots$ be independent and identically distributed random vectors with $\mathbb{E}(X_i) = \mu$ and the covariance matrix $\mathbb{E}(X_i - \mu)(X_i - \mu)^T = \Sigma$. Then,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0, \Sigma).$$

Now, let us explain a quite general theorem that includes the types of results we intended to prove regarding the limiting distribution of $X_n + Y_n$. This result is known as continuous mapping theorem.

Theorem (Continuous mapping theorem)

Let $X_i \in \mathbb{R}^d$ and $X_i \xrightarrow{d} X$, the if $g : \mathbb{R}^d \to \mathbb{R}^\ell$ is a continuous function, then we have $g(X_i) \xrightarrow{d} g(X)$ in distribution.

You will prove a simplified version of this result in the next homework. Similar theorem holds for convergence in probability as well.

Theorem (Continuous mapping theorem)

Let $X_i \in \mathbb{R}^d$ and $X_i \xrightarrow{P} X$, then if $g : \mathbb{R}^d \to \mathbb{R}^\ell$ is a continuous function, then we have $g(X_i) \xrightarrow{P} g(X)$.

As a simple corollary of this theorem you may prove that if random vectors $(X_i, Y_i) \xrightarrow{d} (X, Y)$ in distribution, then $X_i + Y_i \xrightarrow{d} X + Y$. Why?

Applications

Sign test

The sign test was based on the following criterion:

$$\text{reject } H_0 \text{ if } \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq \mu_0) - \frac{1}{2} \right| > \kappa.$$

- We would like to calculate the significance level of the test.
- But this time we assume that the number of samples is high (e.g. $n > 50$). Therefore, we can employ the asymptotic results.

It is straightforward to show that according to CLT, under the null hypothesis we have

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq \mu_0) - \frac{1}{2} \right) \xrightarrow{d} N\left(0, \frac{1}{4}\right).$$

(Please prove it for yourself.) Therefore,

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq \mu_0) - \frac{1}{2} \right| \geq \kappa \right) \approx 2\Phi(-2\kappa\sqrt{n}),$$

where Φ is the CDF of the standard normal distribution, i.e.,

$$\Phi(a) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx.$$

t-distribution

Let $X_1, X_2, \ldots, X_n, \ldots$ be iid with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = 1$.

$$Y_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}}.$$

- If we assume that $X_i \overset{i.i.d.}{\sim} N(0,1)$, then Y_n has a t-distribution with $n-1$ df.
- But what if the distribution is not Gaussian?
- Then the characterization of the limiting distribution is a straightforward task. Here are a few hints:
  - (i) note that $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0,1)$;
  - (ii) prove that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{\mathbf{X}})^2 \xrightarrow{p} 1$;
  - (iii) use this result to prove that $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{\mathbf{X}})^2}$ also converges to 1 in probability?
  - (iv) use Slutsky's theorem to derive the limiting distribution of Y_n.