Convergence of random variables (I)

APM 3F004 EP Asymptotic Statistics

S. Gadat

Ecole Polytechnique

Plan

Introduction

- It is often necessary to consider the distribution of a random variable that is itself a function of several random variables. A simple example is :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- Unfortunately, finding the distribution of $Y = g(X_1, \cdots, X_n)$ exactly is often very difficult even if the joint distribution of the random variables is known exactly. In other cases, we may have only partial information about the joint distribution of $X_1, \cdots, X_n$ in which case it is impossible to determine the distribution of Y.
- However, when n is large, it may be possible to obtain approximations to the distribution of Y even when only partial information about $X_1, \cdots, X_n$ is available ; in many cases, these approximations can be remarkably accurate.

Introduction

- Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with mean μ and variance σ^2 and define

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

to be their sample mean ; we would like to look at the behaviour of the distribution of $\bar{X}_n$ when n is large.

- First of all, it seems reasonable that $\bar{X}_n$ will be close to μ if n is sufficiently large ; that is, the random variable $\bar{X}_n - \mu$ should have a distribution that, for large n, is concentrated around 0 or, more precisely,

$$P\left[\left|\bar{X}_n - \mu\right| \leq \varepsilon\right] \approx 1$$

when ε is small. (Note that $\mathrm{Var}\left(\bar{X}_n\right) = \sigma^2/n \to 0$ as $n \to \infty$.)

Plan

Assumptions

Let (X_n) be a sequence of real-valued random variables and X be a real-valued random variable, all defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
There are mainly four types of convergence involving (X_n) :

- Weak convergence in distribution
- Intermediate convergence in probability
- Convergence in quadratic mean
- Strong convergence almost sure

1 - The weak convergence - "in distribution"

The convergence in distribution of (X_n) is the weakest type of convergence. It only involves the distributions of the variables X_n.

Definition

We say that (X_n) converges in distribution to X, and we denote $X_n \xrightarrow{d} X$ if

$$\lim_{n\to+\infty} \mathbb{E}(\varphi(X_n)) = \mathbb{E}(\varphi(X)), \quad \forall \varphi \in \mathcal{C}_b(\mathbb{R}).$$

Convergence in distribution and $c.d.f.$

The following proposition provides a more practical equivalent statement to the preview definition.

Proposition

(X_n) converges in distribution to X if and only if

$$\lim_{n \to +\infty} F_{X_n}(t) = F_X(t),$$

for any t at which F_X is continuous.

Application

Exercise 1. Let (X_n) be a sequence real-valued random variables such that $X_n \hookrightarrow \mathcal{U}(\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\})$ for any $n \in \mathbb{N}^*$. Find the limit in distribution of (X_n).

Exercise 2. Suppose that $X_1, \cdots, X_n$ are i.i.d. $X_i \sim \mathcal{U}([0, 1])$. Define

$$M_n = \max(X_1, \cdots, X_n)$$

Find the limiting distribution of $n(1 - M_n)$.

Exercise 3. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables $\mathcal{U}(\{0, 1, 2, \ldots, 9\})$ and define

$$U_n = \sum_{k=1}^{n} \frac{X_k}{10^k}$$

Find the limiting distribution of U_n.

Remark

- *A sequence of cumulative distribution functions does not converge necessarily to a cumulative distribution function.*
- *When X_n are continuous, the convergence of the cumulative distribution functions does not imply the convergence of the corresponding $p.d.f$.*
- *A sequence of discrete random variables may converge in distribution to a continuous random variable and vice versa.*
- *When (X_n) converges in distribution to X. This involves the distribution of X rather than X as a random experiment.*

Application

Exercise 4. Let (X_n) be a sequence real-valued random variables such that $X_n \hookrightarrow \mathcal{B}(n, p_n)$. We assume that there is $\lambda \in \mathbb{R}_+$ such that

$$\lim_{n \to +\infty} np_n = \lambda.$$

Show that (X_n) converges in distribution to $\mathcal{P}(\lambda)$.

Remark

In practice, one may consider $\mathcal{P}(np)$ as an approximate law of $\mathcal{B}(n, p)$ whenever $n > 50$ and $np < 10$.

The multivariate convergence in distribution

Definition

Let (X_n) be a sequence of d-dimensional random vectors and let X be a d-dimensional random vector. We say that (X_n) converges in distribution to X if for any $a \in \mathbb{R}^d$,

$$\langle X_n, a \rangle \xrightarrow{d} \langle X, a \rangle.$$

Proposition (Continuous mapping theorem)

Assume that $X_n \xrightarrow{d} X$ and that h is a continuous function, then $h(X_n) \xrightarrow{d} h(X)$.

2 - Intermediate convergence - Convergence "in probability"

Definition

We say that (X_n) converges in probability to X, and we denote $X_n \xrightarrow{\mathbb{P}} X$, if

$$\forall \varepsilon > 0, \lim_{n \to +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

The convergence in probability involves the distribution of $|X_n - X|$.

Exercise

Exercise 5. Suppose that $X_1, \cdots, X_n$ are i.i.d. $X_i \sim \mathcal{U}([0,1])$. Define

$$M_n = \max(X_1, \cdots, X_n)$$

Show that $M_n \xrightarrow{\mathbb{P}} 1$

Exercise 6. Let $Z \hookrightarrow \mathcal{U}((0,1))$ and let $(X_n)_{n\geq 1}$ be such that $X_n \hookrightarrow \mathcal{B}(p_n)$. p_n is determined according to the following rules :

- **Step 1 :** We set $X_1 = 1$.
- **Step 2 :** We set $X_2 = \mathbb{1}_{]0,\frac{1}{2}[}(Z)$ and $X_3 = \mathbb{1}_{[\frac{1}{2},1[}(Z)$.
- **Step 3 :** We set $X_4 = \mathbb{1}_{]0,\frac{1}{4}[}(Z)$, $X_5 = \mathbb{1}_{[\frac{1}{4},\frac{1}{2}[}(Z)$, $X_6 = \mathbb{1}_{[\frac{1}{2},\frac{3}{4}[}(Z)$ and $X_7 = \mathbb{1}_{[\frac{3}{4},1[}(Z)$
- **Step n :** We set $X_n = \mathbb{1}_{[a_n,b_n[}(Z)$

1. Determine p_n.
2. Show that (X_n) converges in probability to $X = 0$.

Convergence in probability implies convergence in distribution

Very important result :

Théorème

The convergence in probability implies the convergence in distribution but the converse is not necessarily true.

Proof : . . .$\square$

The converse implication holds in a very specific case !

Théorème

If $X_n \xrightarrow{d} \delta_c$ *for some* $c \in \mathbb{R}$*, then* $X_n \xrightarrow{\mathbb{P}} X = c$.

The multivariate convergence in probability

Definition

Let (X_n) be a sequence of d-dimensional random vectors and let X be a d-dimensional random vector. We say that (X_n) converges in probability to X if

$$\forall \varepsilon > 0, \lim_{n \to +\infty} \mathbb{P}(\| X_n - X \| > \varepsilon) = 0.$$

Proposition (Continuous mapping theorem)

Assume that $X_n \xrightarrow{\mathbb{P}} X$ and that h is a continuous function, then $h(X_n) \xrightarrow{\mathbb{P}} h(X)$.

3 - Stronger convergence - Convergence "in quadratic mean"

The convergence in quadratic mean is a convergence in the Hilbert space $L^2(\Omega, \mathcal{A}, \mathbb{P})$.

Definition

Assume that for any $n \in \mathbb{N}$, $\mathbb{E}\left(X_n^2\right) < +\infty$ and $\mathbb{E}\left(X^2\right) < +\infty$. We say that (X_n) converges in quadratic mean to X, and we denote $X_n \xrightarrow{q.m.} X$, if

$$\lim_{n \to +\infty} \mathbb{E}\left((X_n - X)^2\right) = 0.$$

Application

Exercise 7. Show that the sequence (X_n) in Exercise 6 converges in quadratic mean.

Two inequalities : Markov and Bienaymé-Tchebychev

Proposition (Markov)

Let X be a non-negative real-valued random variable with $\mathbb{E}(X) < +\infty$. The Markov inequality states that

$$\forall \varepsilon \in \mathbb{R}_+^*, \; \mathbb{P}(X > \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}$$

Proposition (Bienaymé-Tchebychev)

Let X be a real-valued random variable with $\mathbb{V}(X) < +\infty$. The Bienaym-Tchebychev inequality states that

$$\forall \varepsilon \in \mathbb{R}_+^*, \; \mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}$$

Convergence in quadratic mean implies convergence in probability

Proposition

The convergence in quadratic mean implies the convergence in probability but the converse is not necessarily true.

Proof : . . . $\square$

The multivariate convergence in quadratic mean

Definition

Let (X_n) be a sequence of d-dimensional random vectors and let X be a d-dimensional random vector. We say that (X_n) converges in quadratic mean to X if

$$\lim_{n \to +\infty} \mathbb{E}\left(\| X_n - X \|^2\right) = 0.$$

4 - Stronger convergence - almost sure convergence

The almost sure convergence is the most natural and the strongest type of convergence. But it is also the most difficult to prove in practice.

Definition

We say that (X_n) converges to X almost surely, and we denote $X_n \xrightarrow{a.s.} X$, when

$$\mathbb{P}(\{\omega \in \Omega; X_n(\omega) \longrightarrow X(\omega)\}) = 1.$$

The almost sure convergence implies the convergence in probability

Proposition

The almost sure convergence implies the convergence in probability but the converse is not necessarily true.

Application **Exercise 8.** Show that the sequence in exercise 6 does not converge almost surely.

The multivariate almost sure convergence

Definition

Let (X_n) be a sequence of d-dimensional random vectors and let X be a d-dimensional random vector. We say that (X_n) converges almost surely to X if

$$\mathbb{P}\left(\left\{\omega \in \Omega; \lim_{n \to +\infty} \| X_n(\omega) - X(\omega) \| = 0\right\}\right) = 1.$$

5 - Slutsky's Lemma

Very important result in statistics.

Lemme (Slutsky's Lemma)

Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be sequences of random variables defined on a common probability space. Suppose that

$$X_n \xrightarrow{d} X \quad \text{and} \quad Y_n \xrightarrow{\mathbb{P}} c,$$

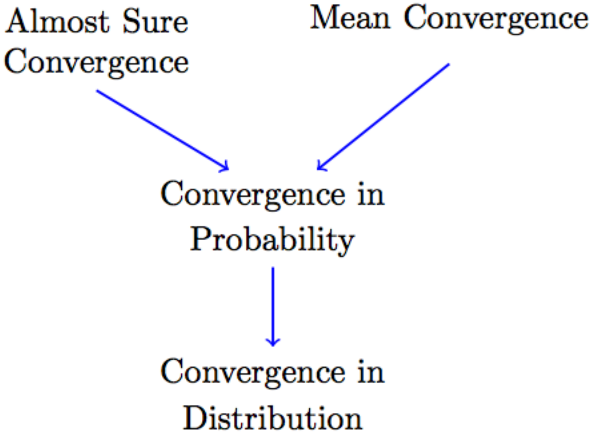
where X is a random variable and $c \in \mathbb{R}$ is a constant. Then :

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

<u>Remark :</u> Generally, used with CLT and LLN (see later on)

Plan

Link between convergences



Combining convergence statements

Remark

You have to be cautious when combining convergence statements. For example :

- *it is possible to add two convergence in probability,*
- *it is possible to add two convergence in quadratic mean*
- *it is possible to add two almost sure convergence*
- *But it does not necessarily work for convergence in distribution.*