Convergence of Random Variables

Saravanan Vijayakumaran
sarva@ee.iitb.ac.in

Department of Electrical Engineering
Indian Institute of Technology Bombay

March 19, 2014

Motivation

Theorem (Weak Law of Large Numbers)

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with finite means μ. Their partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfy

$$\frac{S_n}{n} \xrightarrow{P} \mu \qquad \text{as } n \to \infty.$$

Theorem (Central Limit Theorem)

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with finite means μ and finite non-zero variance σ^2. Their partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfy

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2) \qquad \text{as } n \to \infty.$$

Modes of Convergence

- A sequence of real numbers $\{x_n : n = 1, 2, \ldots\}$ is said to converge to a limit x if for all $\varepsilon > 0$ there exists an $m_\varepsilon \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ for all $n \geq m_\varepsilon$.
- We want to define convergence of random variables but they are functions from Ω to $\mathbb{R}$
- The solution
  - Derive real number sequences from sequences of random variables
  - Define convergence of the latter in terms of the former
- Four ways of defining convergence for random variables
  - Convergence almost surely
  - Convergence in rth mean
  - Convergence in probability
  - Convergence in distribution

Convergence Almost Surely

- Let $X, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$
- For each $\omega \in \Omega$, $X(\omega)$ and $X_n(\omega)$ are reals
- $X_n \to X$ almost surely if $\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}$ is an event whose probability is 1
- "$X_n \to X$ almost surely" is abbreviated as $X_n \xrightarrow{\text{a.s.}} X$

Example

- Let $\Omega = [0, 1]$ and P be the uniform distribution on Ω
- $P(\omega \in [a, b]) = b - a$ for $0 \le a \le b \le 1$
- Let X_n be defined as

$$X_n(\omega) = \begin{cases} n, & \omega \in \left[0, \frac{1}{n}\right) \\ 0, & \omega \in \left[\frac{1}{n}, 1\right] \end{cases}$$

- Let $X(\omega) = 0$ for all $\omega \in [0, 1]$
- $X_n \xrightarrow{\text{a.s.}} X$

Convergence in rth Mean

- Let $X, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$
- Suppose $E[|X^r|] < \infty$ and $E[|X_n^r|] < \infty$ for all n
- $X_n \to X$ in rth mean if

$$E\left(|X_n - X|^r\right) \to 0 \text{ as } n \to \infty$$

  where $r \geq 1$
- "$X_n \to X$ in rth mean" is abbreviated as $X_n \xrightarrow{r} X$
- For $r = 1$, $X_n \xrightarrow{1} X$ is written as "$X_n \to X$ in mean"
- For $r = 2$, $X_n \xrightarrow{2} X$ is written as "$X_n \to X$ in mean square" or $X_n \xrightarrow{\text{m.s.}} X$

Example

- Let $\Omega = [0, 1]$ and P be the uniform distribution on Ω
- Let X_n be defined as

$$X_n(\omega) = \begin{cases} n, & \omega \in \left[0, \frac{1}{n}\right) \\ 0, & \omega \in \left[\frac{1}{n}, 1\right] \end{cases}$$

- Let $X(\omega) = 0$ for all $\omega \in [0, 1]$
- $E[|X_n|] = 1$ and so X_n does not converge in mean to X

Convergence in Probability

- Let $X, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$
- $X_n \to X$ in probability if

$$P\left(|X_n - X| > \epsilon\right) \to 0 \text{ as } n \to \infty \text{ for all } \epsilon > 0$$

- "$X_n \to X$ in probability" is abbreviated as $X_n \xrightarrow{P} X$

Example

- Let $\Omega = [0, 1]$ and P be the uniform distribution on Ω
- Let X_n be defined as

$$X_n(\omega) = \begin{cases} n, & \omega \in \left[0, \frac{1}{n}\right) \\ 0, & \omega \in \left[\frac{1}{n}, 1\right] \end{cases}$$

- Let $X(\omega) = 0$ for all $\omega \in [0, 1]$
- For $\varepsilon > 0$, $P[|X_n - X| > \varepsilon] = P[|X_n| > \varepsilon] \le P[X_n = n] = \frac{1}{n} \to 0$
- $X_n \xrightarrow{P} X$

Convergence in Distribution

- Let $X, X_1, X_2, \ldots$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$
- $X_n \to X$ in distribution if

$$P(X_n \le x) \to P(X \le x) \text{ as } n \to \infty$$

  for all points x where $F_X(x) = P(X \le x)$ is continuous
- "$X_n \to X$ in distribution" is abbreviated as $X_n \xrightarrow{D} X$
- Convergence in distribution is also termed weak convergence

Example

Let X be a Bernoulli RV taking values 0 and 1 with equal probability $\frac{1}{2}$.
Let $X_1, X_2, X_3, \ldots$ be identical random variables given by $X_n = X$ for all n.
The X_n's are not independent but $X_n \xrightarrow{D} X$.
Let $Y = 1 - X$. Then $X_n \xrightarrow{D} Y$.
But $|X_n - Y| = 1$ and the X_n's do not converge to Y in any other mode.

Relations between Modes of Convergence

Theorem

$$
\begin{array}{ccccc}
(X_n \xrightarrow{a.s.} X) & & & & \\
& \searrow & & & \\
& & (X_n \xrightarrow{P} X) & \Rightarrow & (X_n \xrightarrow{D} X) \\
& \nearrow & & & \\
(X_n \xrightarrow{r} X) & & & &
\end{array}
$$

for any $r \geq 1$.

Convergence in Probability Implies Convergence in Distribution

- Suppose $X_n \xrightarrow{P} X$
- Let $F_n(x) = P(X_n \leq x)$ and $F(x) = P(X \leq x)$
- If $\varepsilon > 0$,

$$\begin{aligned}
F_n(x) &= P(X_n \leq x) \\
&= P(X_n \leq x, X \leq x + \varepsilon) + P(X_n \leq x, X > x + \varepsilon) \\
&\leq F(x + \varepsilon) + P(|X_n - X| > \varepsilon) \\
F(x - \varepsilon) &= P(X \leq x - \varepsilon) \\
&= P(X \leq x - \varepsilon, X_n \leq x) + P(X \leq x - \varepsilon, X_n > x) \\
&\leq F_n(x) + P(|X_n - X| > \varepsilon)
\end{aligned}$$

- Combining the above inequalities we have

$$F(x - \varepsilon) - P(|X_n - X| > \varepsilon) \leq F_n(x) \leq F(x + \varepsilon) + P(|X_n - X| > \varepsilon)$$

- If F is continuous at x, $F(x - \varepsilon) \to F(x)$ and $F(x + \varepsilon) \to F(x)$ as $\varepsilon \downarrow 0$
- Since $X_n \xrightarrow{P} X$, $P(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$

Convergence in rth Mean Implies Convergence in Probability

- If $r > s \geq 1$ and $X_n \xrightarrow{r} X$ then $X_n \xrightarrow{s} X$
  - Lyapunov's inequality: If $r > s > 0$, then $\left(E\left[|Y|^s\right]\right)^{\frac{1}{s}} \leq \left(E\left[|Y|^r\right]\right)^{\frac{1}{r}}$
  - If $X_n \xrightarrow{r} X$, then $E\left[|X_n - X|^r\right] \to 0$ and $\left(E\left[|X_n - X|^s\right]\right)^{\frac{1}{s}} \leq \left(E\left[|X_n - X|^r\right]\right)^{\frac{1}{r}}$
- If $X_n \xrightarrow{1} X$ then $X_n \xrightarrow{P} X$
- By Markov's inequality, we have

$$P\left(|X_n - X| > \varepsilon\right) \leq \frac{E\left(|X_n - X|\right)}{\varepsilon}$$

for all $\varepsilon > 0$

Convergence Almost Surely Implies Convergence in Probability

- Let $A_n(\varepsilon) = \{|X_n - X| > \varepsilon\}$ and $B_m(\varepsilon) = \bigcup_{n \geq m} A_n(\varepsilon)$
- $X_n \xrightarrow{\text{a.s.}} X$ if and only if $P\left(B_m(\varepsilon)\right) \to 0$ as $m \to \infty$, for all $\varepsilon > 0$
  - Let

$$\begin{aligned} C &= \{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\} \\ A(\varepsilon) &= \{\omega \in \Omega : \omega \in A_n(\varepsilon) \text{ for infinitely many values of } n\} \\ &= \bigcap_m \bigcup_{n=m}^{\infty} A_n(\varepsilon) \end{aligned}$$

  - $X_n(\omega) \to X(\omega)$ if and only if $\omega \notin A(\varepsilon)$ for all $\varepsilon > 0$
  - $P(C) = 1$ if and only if $P\left(A(\varepsilon)\right) = 0$ for all $\varepsilon > 0$
  - $B_m(\varepsilon)$ is a decreasing sequence of events with limit $A(\varepsilon)$
  - $P(A(\varepsilon)) = 0$ if and only if $P\left(B_m(\varepsilon)\right) \to 0$ as $m \to \infty$
- Since $A_n(\varepsilon) \subseteq B_n(\varepsilon)$, we have $P\left(|X_n - X| > \varepsilon\right) = P\left(A_n(\varepsilon)\right) \to 0$ whenever $P\left(B_n(\varepsilon)\right) \to 0$
- Thus $X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{P} X$

Some Converses

- If $X_n \xrightarrow{D} c$, where c is a constant, then $X_n \xrightarrow{P} c$

$$P\left(|X_n - c| > \varepsilon\right) = P(X_n < c - \varepsilon) + P(X_n > c + \varepsilon) \to 0 \text{ if } X_n \xrightarrow{D} c$$

- If $P_n(\varepsilon) = P\left(|X_n - X| > \varepsilon\right)$ satisfies $\sum_n P_n(\varepsilon) < \infty$ for all $\varepsilon > 0$, then $X_n \xrightarrow{\text{a.s.}} X$
  - Let $A_n(\varepsilon) = \{|X_n - X| > \varepsilon\}$ and $B_m(\varepsilon) = \bigcup_{n \geq m} A_n(\varepsilon)$

$$P\left(B_m(\varepsilon)\right) \leq \sum_{n=m}^{\infty} P\left(A_n(\varepsilon)\right) = \sum_{n=m}^{\infty} P_n(\varepsilon) \to 0 \text{ as } m \to \infty$$

  - $X_n \xrightarrow{\text{a.s.}} X$ if and only $P\left(B_m(\varepsilon)\right) \to 0$ as $m \to \infty$, for all $\varepsilon > 0$

Borel-Cantelli Lemmas

- Let $A_1, A_2, \ldots$ be an infinite sequence of events from $(\Omega, \mathcal{F}, P)$
- Consider the event that infinitely many of the A_n occur

$$A = \{A_n \text{ i.o.}\} = \bigcap_n \bigcup_{m=n}^{\infty} A_m$$

Theorem

Let A be the event that infinitely many of the A_n occur. Then

- $P(A) = 0$ *if* $\sum_n P(A_n) < \infty$*,*
- $P(A) = 1$ *if* $\sum_n P(A_n) = \infty$ *and* $A_1, A_2, A_3, \ldots$ *are independent events*

Proof of first lemma.

We have $A \subseteq \bigcup_{m=n}^{\infty} A_m$ for all n

$$P(A) \leq \sum_{m=n}^{\infty} P(A_m) \to 0 \text{ as } n \to 0$$

□

Proof of Second Borel-Cantelli Lemma

$$A^c = \bigcup_n \bigcap_{m=n}^{\infty} A_m^c$$

$$\begin{aligned} P\left(\bigcap_{m=n}^{\infty} A_m^c\right) &= \lim_{r\to\infty} P\left(\bigcap_{m=n}^{r} A_m^c\right) = \lim_{r\to\infty} \prod_{m=n}^{r} [1 - P(A_m)] = \prod_{m=n}^{\infty} [1 - P(A_m)] \\ &\leq \prod_{m=n}^{\infty} \exp\left[-P(A_m)\right] = \exp\left(-\sum_{m=n}^{\infty} P(A_m)\right) = 0 \end{aligned}$$

Thus

$$P(A^c) = \lim_{n\to\infty} P\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0$$

Reference

- Chapter 7, *Probability and Random Processes*, Grimmett and Stirzaker, Third Edition, 2001.