Modes of convergence for random variables

Samy Tindel

Purdue University

Probability Theory 1 - MA 538

Mostly taken from
Probability and Random Processes (Sections 7.1-7.2)
by Grimmett-Stirzaker

Outline

Outline

Outline

Outline

Probability space

Probability space: $(\Omega, \mathcal{F}, \mathbf{P})$ with

- Ω set
- $\mathcal{F}$ a σ-algebra
- $\mathbf{P}$ probability measure

Complete probability space

Hypothesis: We assume that $\mathbf{P}$ is complete, i.e

$$A \in \mathcal{F} \text{ such that } \mathbf{P}(A) = 0, \text{ and } B \subset A$$
$$\Longrightarrow$$
$$B \in \mathcal{F} \text{ and } \mathbf{P}(B) = 0.$$

Remark: A probability can always be completed

Simple examples (1)

Tossing 2 dice:

- $\Omega = \{1, 2, 3, 4, 5, 6\}^2$
- $\mathcal{F} = \mathcal{P}(\Omega)$
- $\mathbf{P}(A) = \frac{|A|}{36}$

Uniform distribution on $[0, 1]$:

- $\Omega = [0, 1]$
- $\mathcal{F} = \mathcal{B}([0, 1])$
- $\mathbf{P} = \lambda$, Lebesgue measure

Simple examples (2)

Gaussian law on $\mathbb{R}$:

- $\Omega = \mathbb{R}$
- $\mathcal{F} = \mathcal{B}(\mathbb{R})$
- $\mathbf{P}(A) = \frac{1}{(2\pi)^{1/2}} \int_A e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx$, for $A \in \mathcal{F}$

Typical example for this course

Proposition 1.

Let $\Omega = \ell^p$ with $p \in (1, \infty)$. We set:

$$d(u, v) = \left(\sum_{n \geq 1} |u_n - v_n|^p \right)^{1/p} .$$

Then Ω is a complete metric separable space.

Random variables

Definition 2.

Let

- $(\Omega, \mathcal{F}, \mathbf{P})$ complete probability space
- A function $X : \Omega \to \mathbb{R}$

Then

X is said to be a random variable if X is measurable

Independence (1)

Independence of r.v: Let $(X_j)_{j\in J}$ r.v in $\mathbb{R}^n$.
Those r.v are said to be independent if for all $m \geq 2$:

- For every $j_1, \ldots, j_m \in J$, the r.v $(X_{j_1}, \ldots, X_{j_m})$ are $\perp\!\!\!\perp$
- Otherwise stated: for all $A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R}^n)$ we have

$$\mathbf{P}\left(X_{j_1} \in A_1, \ldots, X_{j_m} \in A_m\right) = \prod_{k=1}^{m} \mathbf{P}\left(X_{j_k} \in A_k\right)$$

Independence (2)

Independence of σ-algebras: Let $(\mathcal{F}_j)_{j\in J}$ σ-algebras, $\mathcal{F}_j \subset \mathcal{F}$.
Those σ-algebras are said to be independent if for all $m \geq 2$:

- For all $j_1, \ldots, j_m \in J$, the σ-algebras $(\mathcal{F}_{j_1}, \ldots, \mathcal{F}_{j_m})$ are $\perp\!\!\!\perp$
- Otherwise stated: for all $B_1 \in \mathcal{F}_{j_1}, \ldots, B_m \in \mathcal{F}_{j_m}$ we have

$$\mathbf{P}\left(\bigcap_{k=1}^{m} B_k\right) = \prod_{k=1}^{m} \mathbf{P}\left(B_k\right)$$

π-systems and λ-systems

π-system: Let $\mathcal{P}$ family of subsets of Ω. $\mathcal{P}$ is a π-system if:

$$A, B \in \mathcal{P} \quad \Longrightarrow \quad A \cap B \in \mathcal{P}$$

λ-system: Let $\mathcal{L}$ family of subsets of Ω. $\mathcal{L}$ is a λ-system if:

1. $\Omega \in \mathcal{L}$
2. If $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$
3. If for $j \geq 1$ we have:
   - $A_j \in \mathcal{L}$
   - $A_j \cap A_i = \emptyset$ if $j \neq i$

   Then $\cup_{j\geq 1} A_j \in \mathcal{L}$

Dynkin's π-λ lemma

Proposition 3.

Let $\mathcal{P}$ et $\mathcal{L}$ such that:

- $\mathcal{P}$ is a π-system
- $\mathcal{L}$ is a λ-system
- $\mathcal{P} \subset \mathcal{L}$

Then $\sigma(\mathcal{P}) \subset \mathcal{L}$

Application of Dynkin's π-λ lemma

Proposition 4.

Let:

- $X_1, \ldots, X_n$ r.v with values in $\mathbb{R}^m$.
- $X \equiv (X_1, \ldots, X_n) \in \mathbb{R}^{m \times n}$.
- $\mu_{X_j} = \mathcal{L}(X_j)$ and $\mu_X = \mathcal{L}(X)$.

Then the two following assertions are equivalent:

1. $X_1, \ldots, X_n$ are independent
2. $\mu_X = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$ on $\mathcal{B}(\mathbb{R}^{m \times n})$

Proof (1)

Definition of two systems: We set

$$\mu_1 = \mu_X, \quad \text{and} \quad \mu_2 = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n},$$

and

$$\begin{aligned}
\mathcal{P} &\equiv \left\{A \in \mathcal{B}(\mathbb{R}^{m\times n});\, A = A_1 \times \cdots \times A_n, \text{ where } A_j \in \mathcal{B}(\mathbb{R}^m)\right\} \\
\mathcal{L} &\equiv \left\{B \in \mathcal{B}(\mathbb{R}^{m\times n});\, \mu_1(B) = \mu_2(B)\right\}.
\end{aligned}$$

Proof (2)

Application of Dynkin's lemma: We have

- $\mathcal{P}$ is a π-system
- $\mathcal{L}$ is a λ-system
- $\mu_1(C) = \mu_2(C)$ for all $C \in \mathcal{P}$

Thus $\sigma(\mathcal{P}) \subset \mathcal{L}$, and $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R}^{m \times n})$

Conclusion:

$$\mu_1(A) = \mu_2(A) \text{ for all } A \in \mathcal{B}(\mathbb{R}^{m \times n})$$

Outline

Experiment

Procedure:

- Consider a plane ruled by lines $y = k$, with $k \in \mathbb{Z}$
- Take a needle with length 1
- Fling the needle n times on the plane

Outcome: We record, for $i = 1, \ldots, n$,

- $X_i \equiv \mathbf{1}_{A_i}$, where $A_i = (i$-th needle intersect a line$)$
- $S_n \equiv \#$ times the needle intersects the line

Simulation:

This website from UIUC

Limiting result

Proposition 5.

Under the above conditions we have

$$\mathbf{P}(A_i) = \frac{2}{\pi}$$

$$\frac{S_n}{n} \longrightarrow \frac{2}{\pi}$$

Bernoulli random variable (1)

Notation:

$$X \sim \mathcal{B}(p) \text{ with } p \in (0,1)$$

State space:

$$\{0,1\}$$

Pmf:

$$\mathbf{P}(X=0) = 1-p, \quad \mathbf{P}(X=1) = p$$

Expected value, variance, generating function:

$$\mathbf{E}[X] = p, \qquad \mathbf{Var}(X) = p(1-p), \qquad G_X(s) = (1-p) + p\,s$$

Bernoulli random variable (2)

Use 1, success in a binary game:

- Example 1: coin tossing
  - $X = 1$ if H, $X = 0$ if T
  - We get $X \sim \mathcal{B}(1/2)$
- Example 2: dice rolling
  - $X = 1$ if outcome $= 3$, $X = 0$ otherwise
  - We get $X \sim \mathcal{B}(1/6)$

Use 2, answer yes/no in a poll

- $X = 1$ if a person feels optimistic about the future
- $X = 0$ otherwise
- We get $X \sim \mathcal{B}(p)$, with unknown p

Jacob Bernoulli

Some facts about Bernoulli:

- Lifespan: 1654-1705, in Switzerland
- Discovers constant e
- Establishes divergence of $\sum \frac{1}{n}$
- Contributions in diff. eq
- First law of large numbers
- Bernoulli:
  family of 8 prominent mathematicians
- Fierce math fights between brothers

Binomial random variable (1)

Notation:

$$X \sim \text{Bin}(n, p), \text{ for } n \geq 1,\ p \in (0, 1)$$

State space:

$$\{0, 1, \ldots, n\}$$

Pmf:

$$\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n$$

Expected value, variance and generating function:

$$\mathbf{E}[X] = np, \qquad \mathbf{Var}(X) = np(1-p), \qquad G_X(s) = [(1-p) + p\,s]^n$$

Binomial random variable (2)

Use 1, Number of successes in a Bernoulli trial:

- Example: Roll a dice 9 times.
- $X = \#$ of 3 obtained
- We get $X \sim \text{Bin}(9, 1/6)$
- $\mathbf{P}(X = 2) = 0.28$

Use 2: Counting a feature in a repeated trial:

- Example: stock of 1000 pants with 10% defects
- Draw 15 times a pant at random
- $X = \#$ of pants with a defect
- We get $X \sim \text{Bin}(15, 1/10)$

Binomial random variable (3)

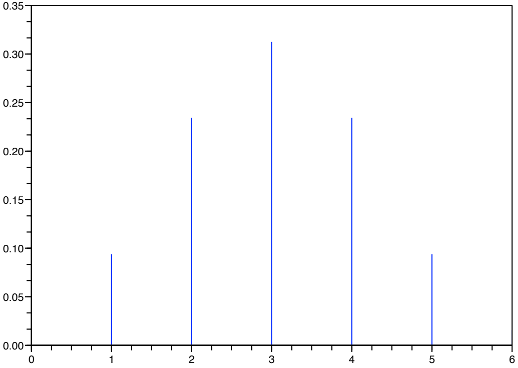


Figure: Pmf for $\text{Bin}(6; 0.5)$. x-axis: k. y-axis: $\mathbf{P}(X = k)$

Binomial random variable (4)

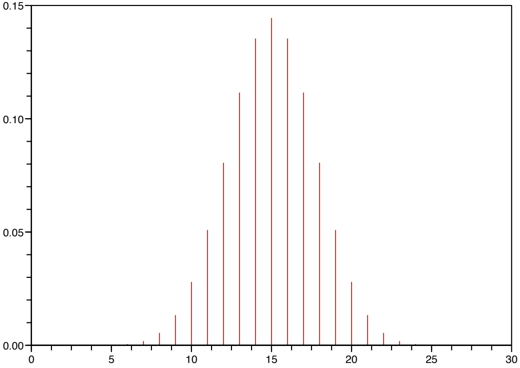


Figure: Pmf for $\text{Bin}(30; 0.5)$. x-axis: k. y-axis: $\mathbf{P}(X = k)$

Uniform random variable (1)

Notation:

$$X \sim \mathcal{U}([\alpha, \beta]), \text{ with } \alpha < \beta$$

State space:

$$[\alpha, \beta]$$

Density:

$$f(x) = \frac{1}{\beta - \alpha} \mathbf{1}_{[\alpha,\beta]}(x)$$

Expected value and variance:

$$\mathbf{E}[X] = \frac{\alpha + \beta}{2}, \qquad \mathbf{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

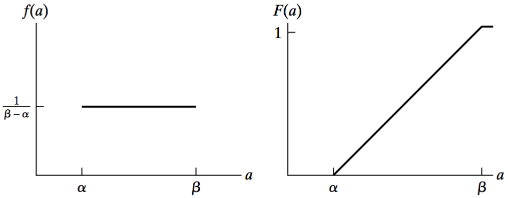
Uniform random variable (2)

Use:

- $\mathcal{U}([0,1])$ only r.v directly accessible on a computer
  $\hookrightarrow$ `rand` function

Example of computation: if $X \sim \mathcal{U}([8,10])$, then

$$\mathbf{P}(7.5 < X < 9.5) = \frac{1}{2}\int_{8}^{9.5} dx = \frac{9.5-8}{2} = \frac{3}{4}$$



Experiment (repeated)

Procedure:

- Consider a plane ruled by lines $y = k$, with $k \in \mathbb{Z}$
- Take a needle with length 1
- Fling the needle n times on the plane

Outcome: We record, for $i = 1, \ldots, n$,

- $X_i \equiv \mathbf{1}_{A_i}$, where $A_i = (i\text{-th needle intersects a line})$
- $S_n \equiv \#$ times the needle intersects the line

Proof of Proposition 5 (1)

Notation: We define

- $(X_i, Y_i) \equiv$ Coordinates of the center of the i-th needle
- $\Theta_i \equiv$ angle (i-th needle, x-axis)
- $Z_i = d\left((X_i, Y_i),\ \text{nearest line underneath}\right) = Y_i - \lfloor Y_i \rfloor$

Model: We assume

1. $Z_i \sim \mathcal{U}([0,1])$
2. $\Theta_i \sim \mathcal{U}([0,\pi])$
3. $Z_i \perp\!\!\!\perp \Theta_i$
4. $\{Z_i;\ i \geq 1\}$ i.i.d sequence
5. $\{\Theta_i;\ i \geq 1\}$ i.i.d sequence

Proof of Proposition 5 (2)

Expression for A_i: We have

$$A_i = A_i^- \cup A_i^+$$

with

$$\begin{aligned} A_i^- &= \left\{ Z_i \le \frac{1}{2}, \text{ and } Z_i < \frac{1}{2}\sin(\Theta_i) \right\} \\ A_i^+ &= \left\{ Z_i > \frac{1}{2}, \text{ and } 1 - Z_i < \frac{1}{2}\sin(\Theta_i) \right\} \end{aligned}$$

Proof of Proposition 5 (3)

Computing $\mathbf{P}(A_i)$: We write

$$\begin{aligned}\mathbf{P}(A_i) &= \mathbf{P}(A_i^-) + \mathbf{P}(A_i^+) \\ &= 2\mathbf{P}(A_i^-) \\ &= \frac{2}{\pi}\int_0^\pi d\theta \int_0^{\frac{1}{2}\sin(\theta)} dz\end{aligned}$$

Thus

$$\mathbf{P}(A_i) = \frac{2}{\pi}$$

Proof of Proposition 5 (4)

Some laws: We have

$$X_i \sim \mathcal{B}\left(\frac{2}{\pi}\right)$$

$$S_n \sim \text{Bin}\left(n, \frac{2}{\pi}\right)$$

Limit: By De Moivre,

$$\frac{S_n}{n} \longrightarrow \frac{2}{\pi}$$

Outline

Aim of this chapter

Problem with limit statement:

- For every $n \geq 1$, we have $S_n : \Omega \to \mathbb{R}$
- S_n is thus a function
- We don't know exactly what $\frac{S_n}{n} \longrightarrow \frac{2}{\pi}$ means!

Aim of this chapter:

- Explore different modes of convergence for random variables

Preliminary step:

- Explore different modes of convergence for functions

Setting for convergence of functions

Sequence of functions: We consider

- A sequence $\{f_n;\ n \geq 1\}$ with

$$f_n : [0, 1] \longrightarrow \mathbb{R}$$

Aim of subsection: Review modes for

$$\lim_{n\to\infty} f_n$$

Pointwise convergence

Definition 6.

Let

- $\{f_n;\ n \geq 1\}$ sequence of measurable functions

We assume

$$\lim_{n\to\infty} f_n(x) = f(x)\,, \text{ for all } x \in [0,1]$$

Then we say that

$$f_n \longrightarrow f \text{ pointwise}$$

Almost everywhere convergence

Definition 7.

Let

- $\{f_n;\ n \geq 1\}$ sequence of measurable functions

We assume

$$\lim_{n \to \infty} f_n(x) = f(x)\,, \text{ for almost every } x \in [0,1]$$

Then we say that

$$f_n \longrightarrow f \text{ almost everywhere}$$

L^p convergence

Definition 8.

Let

- $\{f_n;\ n \geq 1\}$ sequence of measurable functions

We assume

$$\lim_{n\to\infty} \|f_n - f\|_{L^p([0,1])} = 0$$

Then we say that

$$f_n \longrightarrow f \text{ in } L^p([0,1])$$

Convergence in measure

Definition 9.

Let

- $\{f_n;\ n \geq 1\}$ sequence of measurable functions

We assume that for all $\varepsilon > 0$

$$\lim_{n\to\infty} \lambda\left(\{u \in [0,1];\ |f_n(u) - f(u)| > \varepsilon\}\right) = 0$$

Then we say that

$$f_n \longrightarrow f \text{ in measure}$$

Relations between convergences (1)

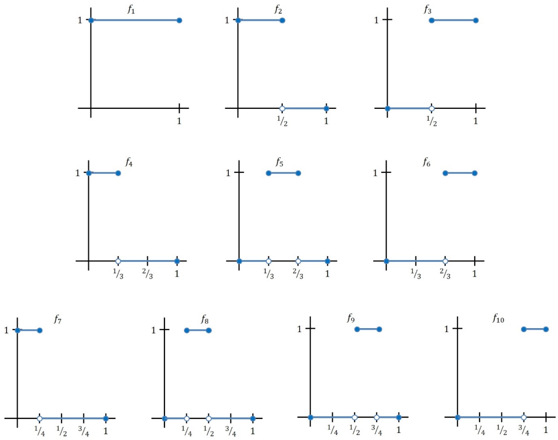
Examples of relations for functions on $[0, 1]$:

- $f_n(x) = x^n$
  $\hookrightarrow$ converges almost everywhere, not pointwise
- $g_n(x) = n\mathbf{1}_{[0,1/n]}(x)$
  $\hookrightarrow$ converges almost everywhere, not in L^1

Relations between convergences (2)

Another example of relation for functions on $[0,1]$:

- $h_n = \mathbf{1}_{[0,1]},\ \mathbf{1}_{[0,1/2]}\ ,\ \mathbf{1}_{[1/2,1]},\ \mathbf{1}_{[0,1/3]},\ \mathbf{1}_{[1/3,2/3]},\ldots$
  $\hookrightarrow$ converges in measure, not almost everywhere



Outline

Outline

Almost sure convergence

Definition 10.

Let

- $\{X_n;\, n \geq 1\}$ sequence of random variables on $(\Omega, \mathcal{F}, \mathbf{P})$
- Another random variable X defined on $(\Omega, \mathcal{F}, \mathbf{P})$

We assume

$$\mathbf{P}\left(\left\{\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Then we say that

$$X_n \longrightarrow X \text{ almost surely}$$

Convergence in L^p

Definition 11.

Let

- $\{X_n;\, n \geq 1\}$ sequence of r.v in $L^r(\Omega)$
- Another random variable $X \in L^r(\Omega)$

We assume

$$\lim_{n\to\infty} \mathbf{E}\left[|X_n - X|^r\right] = 0.$$

Then we say that

$$X_n \longrightarrow X \text{ in } L^r(\Omega) \text{ (or in } r\text{-th mean)}$$

Convergence in probability

Definition 12.

Let

- $\{X_n;\, n \geq 1\}$ sequence of random variables on $(\Omega, \mathcal{F}, \mathbf{P})$
- Another random variable X defined on $(\Omega, \mathcal{F}, \mathbf{P})$

We assume that for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(|X_n - X| > \varepsilon\right) = 0.$$

Then we say that

$$X_n \longrightarrow X \text{ in probability}$$

Convergence in distribution

Definition 13.

Let

- $\{X_n;\ n \geq 1\}$ sequence of random variables on $(\Omega, \mathcal{F}, \mathbf{P})$
- Another random variable X defined on $(\Omega, \mathcal{F}, \mathbf{P})$

We assume that for all points $x \in \mathbb{R}$ such that F_X is continuous,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Then we say that

$$X_n \longrightarrow X \text{ in distribution}$$

Remarks about convergence in distribution

1. The central limit theorem
   $\hookrightarrow$ is a convergence in distribution
2. Ergodic theorems for Markov chains
   $\hookrightarrow$ are convergences in distributions
3. Convergence in distribution
   $\hookrightarrow$ does not refer to a specific $(\Omega, \mathcal{F}, \mathbf{P})$

A Bernoulli example

A Bernoulli sequence: We consider

- $X \sim \mathcal{B}(1/2)$
- $X_n = X$ for all $n \geq 1$
- $Y = 1 - X$

Convergences:

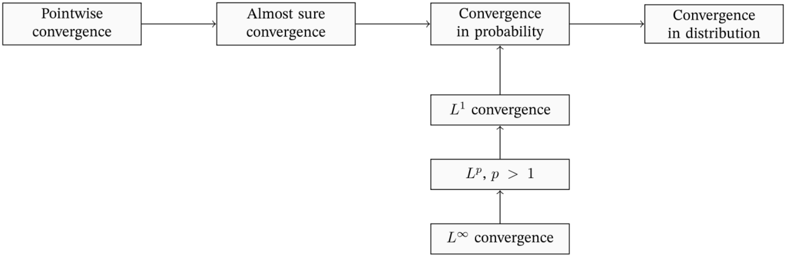
1. We have

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

2. X_n does not converge to X in any other mode

Relations between modes of convergence

Theorem 14.



Outline

Convergence in probability and in distribution

Proposition 15.

Let

- X_n sequence of random variables
- Assume $X_n \xrightarrow{\mathrm{P}} X$

Then

$$X_n \xrightarrow{(\mathrm{d})} X$$

Proof of Proposition 15 (1)

Notation: Set

$$F_n(x) = \mathbf{P}(X_n \le x), \qquad F(x) = \mathbf{P}(X \le x)$$

Aim: Prove that

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ if } F \text{ is continuous at } x$$

Proof of Proposition 15 (2)

1st decomposition: We have

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

2nd decomposition: We have

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

Summary:

$$F(x-\varepsilon) - \mathbf{P}\left(|X_n - X| > \varepsilon\right) \le F_n(x) \le F(x+\varepsilon) + \mathbf{P}\left(|X_n - X| > \varepsilon\right)$$

Proof of Proposition 15 (3)

Limits as $n \to \infty$: Since $X_n \xrightarrow{(P)} X$, we have

$$F(x-\varepsilon) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x+\varepsilon)$$

Limits as $\varepsilon \to 0$: If F is continuous at x, we get

$$F(x) = \liminf_{n\to\infty} F_n(x) = \limsup_{n\to\infty} F_n(x) = F(x)$$

Convergence in $L^p(\Omega)$

Proposition 16.

Let

- X_n sequence of random variables
- Assume $X_n \xrightarrow{L^s} X$ for $s > r$

Then

$$X_n \xrightarrow{L^r} X$$

Proof of Proposition 16

Inequality on norms: We have

$$\|X_n - X\|_r \leq \|X_n - X\|_s$$

Counter-example

Definition of a sequence: We consider independent r.v with

$$\mathbf{P}(X_n = n) = \frac{1}{n^{\frac{1}{2}(r+s)}}, \qquad \mathbf{P}(X_n = 0) = 1 - \frac{1}{n^{\frac{1}{2}(r+s)}}$$

Convergence: If $r < s$ we have

1. $X_n \xrightarrow{L^r} 0$
2. X_n does not converge in L^s

Markov's inequality

Proposition 17.

Let

$$X \text{ random variable with } X \in L^1(\Omega)$$

Then for all $a > 0$ we have

$$\mathbf{P}\left(|X| \geq a\right) \leq \frac{\mathbf{E}\left[|X|\right]}{a}$$

Andrey Markov

Andrey Markov's life:

- Lifespan: 1856-1922, $\simeq$ St Petersburg
- Not a very good student
  $\hookrightarrow$ except in math
- Contributions in analysis and probability
- Used chains for
  $\hookrightarrow$ appearance of vowels
- Professor in St Petersburg
  - Suspended after 1908 students riots
  - Resumed teaching in 1917

Fact: More than 50 mathematical objects named after Markov!!

Proof of Proposition 17

Deterministic inequality: Set

$$A = \{|X| \geq a\}$$

Then we have

$$|X| \geq a\,\mathbf{1}_A\,, \quad \text{almost surely}$$

Expectations: Taking expectations above, we get

$$\mathbf{E}\left[|X|\right] \geq a\,\mathbf{P}(A)$$

Convergence in $L^p(\Omega)$ and in probablity

Proposition 18.

Let

- X_n sequence of random variables
- Assume $X_n \xrightarrow{L^1} X$

Then

$$X_n \xrightarrow{P} X$$

Proof of Proposition 18

Applying Markov's inequality: For $\varepsilon > 0$, we have

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

Then take $n \to \infty$

Counter-example

Definition of a sequence: We consider independent r.v with

$$\mathbf{P}\left(X_n = n^3\right) = \frac{1}{n^2}, \qquad \mathbf{P}\left(X_n = 0\right) = 1 - \frac{1}{n^2}$$

Convergence: We have

1. $X_n \xrightarrow{P} 0$
2. X_n does not converge in L^1

Proof of counter-example for X_n (1)

Some notation: For $\varepsilon > 0$ and $X = 0$ set:

$$A_k(\varepsilon) = \{|X_k - X| > \varepsilon\}$$

Convergence in probability: We have

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}\left(A_n(\varepsilon)\right) &= \lim_{n\to\infty} \mathbf{P}\left(X_n = n^3\right) \\ &= \lim_{n\to\infty} \frac{1}{n^2} \\ &= 0\end{aligned}$$

Thus

$$X_n \xrightarrow{\mathrm{P}} 0$$

Proof of counter-example for X_n (2)

Non convergence in L^1: We have

$$\mathbf{E}[|X_n|] = \mathbf{E}[X_n] = n$$

Thus

$$X_n \overset{L^1}{\not\to} 0$$

Outline

Limsup of sets

Definition 19.

Let

- $\{A_n;\, n \geq 1\}$ sequence in $\mathcal{F}$

We define

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

Interpretation: We also have

$$\limsup_{n\to\infty} A_n = \{\omega \in \Omega;\, \omega \text{ belongs to an infinity of } A_n\text{'s}\}$$

Borel-Cantelli lemma

Theorem 20.

Let

- $\{A_n;\, n \geq 1\}$ sequence in $\mathcal{F}$

We assume

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$$

Then we have

$$\mathbf{P}\left(\limsup_{n\to\infty} A_n\right) = 0$$

Emile Borel

Emile Borel's life:

- Lifespan: 1872-1956, $\simeq$ Paris
- $\#$ 1 student in France
  $\hookrightarrow$ for his academic year
- Contributions in analysis and probability
- Active in politics
- Minister of Navy in 1924-25
- Resistance against nazi occupation
- Introduced the ∞ monkey theorem

Fact: "Only" 14 mathematical objects named after Borel ...

Proof of Theorem 20 (1)

A non-increasing sequence: For $N \geq 1$ define

$$B_N = \bigcup_{k=N}^{\infty} A_k$$

Then

1. $N \mapsto B_N$ is non-increasing
2. $\limsup_{n\to\infty} A_n = \bigcap_{N=1}^{\infty} B_N$

Proof of Theorem 20 (2)

Computing the probability: We have

$$\begin{aligned}
\mathbf{P}\left(\limsup_{n\to\infty} A_n\right) &= \mathbf{P}\left(\bigcap_{N=1}^{\infty} B_N\right) \\
&= \lim_{N\to\infty} \mathbf{P}\left(B_N\right) \\
&= \lim_{N\to\infty} \mathbf{P}\left(\bigcup_{k=N}^{\infty} A_k\right) \\
&\leq \lim_{N\to\infty} \sum_{k=N}^{\infty} \mathbf{P}\left(A_k\right) \\
&= 0
\end{aligned}$$

A.s convergence and limsup

Proposition 21.

Consider

- $\{X_n;\ n \geq 1\}$ sequence of random variables
- For $\varepsilon > 0$ set $A_n(\varepsilon) = \{|X_n - X| > \varepsilon\}$

Then

1. We have

$$X_n \xrightarrow{\text{a.s}} X \iff \mathbf{P}\left(\limsup_{n\to\infty} A_n(\varepsilon)\right) = 0 \text{ for all } \varepsilon > 0$$

2. It holds:

$$\sum_{n=1}^{\infty} \mathbf{P}\left(A_n(\varepsilon)\right) < \infty \text{ for all } \varepsilon > 0 \implies X_n \xrightarrow{\text{a.s}} X$$

Proof of Proposition 21 (1)

Claim: Let

$$
\begin{aligned}
C &= \left\{\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} \\
A(\varepsilon) &= \limsup_{n\to\infty} A_n(\varepsilon)
\end{aligned}
$$

Then we have

$$
C = \bigcap_{\varepsilon>0} (A(\varepsilon))^c = \bigcap_{m\geq 1} \left(A\left(\frac{1}{m}\right)\right)^c
$$

Proof of Proposition 21 (2)

Application for almost sure convergence: We have

$$\begin{aligned}
\mathbf{P}\left(C^c\right) = 0 \iff & \mathbf{P}\left(\bigcup_{m\geq 1} A\left(\frac{1}{m}\right)\right) = 0 \\
\iff & \lim_{n\to\infty} \mathbf{P}\left(A\left(\frac{1}{m}\right)\right) = 0 \\
\iff & \mathbf{P}\left(A\left(\frac{1}{m}\right)\right) = 0, \quad \text{for all } m \geq 1
\end{aligned}$$

This proves item 1

Proof of Proposition 21 (3)

Proof of item 2: We write

$$\sum_{n=1}^{\infty} \mathbf{P}\left(A_n(\varepsilon)\right) < \infty \text{ for all } \varepsilon > 0$$

$$\Longrightarrow \mathbf{P}\left(\limsup_{n\to\infty} A_n(\varepsilon)\right) = 0 \text{ for all } \varepsilon > 0$$

$$\Longrightarrow \mathbf{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

A.s convergence and convergence in probability

Proposition 22.

Consider

- $\{X_n;\, n \geq 1\}$ sequence of random variables

Then we have:

$$X_n \xrightarrow{\text{a.s}} X \quad \Longrightarrow \quad X_n \xrightarrow{\text{P}} X$$

Proof of Proposition 22

Defining some sets: For $\varepsilon > 0$ set:

$$\begin{aligned} A_k(\varepsilon) &= \{|X_k - X| > \varepsilon\} \\ B_n(\varepsilon) &= \bigcup_{k \geq n} A_k(\varepsilon) \end{aligned}$$

Evaluating some probabilities: We have seen

$$X_n \xrightarrow{\text{a.s}} X \quad \Longrightarrow \quad \lim_{n\to\infty} \mathbf{P}\left(B_n(\varepsilon)\right) = 0$$

Thus

$$\lim_{n\to\infty} \mathbf{P}\left(A_n(\varepsilon)\right) = 0, \text{ for all } \varepsilon > 0$$

Counter-example

Definition of a sequence: We consider independent r.v with

$$X_n \sim \mathcal{B}\left(\frac{1}{n}\right)$$

Convergence: We have

1. $X_n \xrightarrow{\text{P}} 0$
2. X_n does not converge almost surely

Proof of counter-example

Recalling notation: For $\varepsilon > 0$ set:

$$\begin{aligned} A_k(\varepsilon) &= \{|X_k - X| > \varepsilon\} \\ B_n(\varepsilon) &= \bigcup_{k \geq n} A_k(\varepsilon) \end{aligned}$$

Convergence in probability: We have

$$\begin{aligned} \lim_{n\to\infty} \mathbf{P}\left(A_n(\varepsilon)\right) &= \lim_{n\to\infty} \mathbf{P}\left(X_n = 1\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \\ &= 0 \end{aligned}$$

Thus

$$X_n \xrightarrow{\mathrm{P}} 0$$

Proof of counter-example

Almost sure convergence: We have

$$
\begin{aligned}
\mathbf{P}\left(B_n(\varepsilon)\right) &= 1 - \mathbf{P}\left((B_n(\varepsilon))^c\right) \\
&= 1 - \mathbf{P}\left(\bigcap_{k \geq n} (A_k(\varepsilon))^c\right) \\
&= 1 - \mathbf{P}\left(\bigcap_{k \geq n} (X_k = 0)\right) \\
&= 1 - \prod_{k=n}^{\infty}\left(1 - \frac{1}{k}\right) \\
&= 1
\end{aligned}
$$

Thus

$$X_n \overset{\text{a.s}}{\not\to} 0$$

Non comparison between a.s and L^1-convergence

Proposition 23.

One can find

1. $\{X_n;\, n \geq 1\}$ sequence of random variables such that
$$X_n \xrightarrow{\text{a.s}} X\,, \quad \text{but} \quad X_n \overset{L^1}{\not\to} X$$
2. $\{Y_n;\, n \geq 1\}$ sequence of random variables such that
$$Y_n \xrightarrow{L^1} Y\,, \quad \text{but} \quad Y_n \overset{\text{a.s}}{\not\to} Y$$

Proof of counter-example for X_n (1)

Definition of a sequence (repeated):
We consider independent r.v with

$$\mathbf{P}\left(X_n = n^3\right) = \frac{1}{n^2}, \qquad \mathbf{P}\left(X_n = 0\right) = 1 - \frac{1}{n^2}$$

Convergence: We have

1. $X_n \xrightarrow{\text{a.s}} 0$
2. X_n does not converge in L^1

Proof of counter-example for X_n (2)

Some notation: For $\varepsilon > 0$ set:

$$A_k(\varepsilon) = \{|X_k - X| > \varepsilon\}$$

Almost sure convergence: We have

$$\sum_{n=1}^{\infty} \mathbf{P}\left(A_n(\varepsilon)\right) = \sum_{n=1}^{\infty} \mathbf{P}\left(X_n = n^3\right) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

Thus

$$X_n \xrightarrow{\text{a.s}} 0$$

Proof of counter-example for X_n (3)

Non convergence in $L^1(\Omega)$: We have already seen that

$$\mathbf{E}[|X_n|] = \mathbf{E}[X_n] = n$$

Thus

$$X_n \overset{L^1}{\not\to} 0$$

Outline

Case for which $\xrightarrow{(d)}$ yields $\xrightarrow{P}$

Proposition 24.

Consider

- $\{X_n;\, n \geq 1\}$ sequence of random variables

Assume

$$X_n \xrightarrow{(d)} c, \text{ where } c \text{ is a constant}$$

Then we have:

$$X_n \xrightarrow{P} c$$

Proof of Proposition 24

Expression in terms of cdf: We have

$$\begin{aligned}\mathbf{P}\left(|X_n - c| > \varepsilon\right) &= \mathbf{P}\left(X_n < c - \varepsilon\right) + \mathbf{P}\left(X_n > c + \varepsilon\right) \\ &= \mathbf{P}\left(X_n < c - \varepsilon\right) + 1 - \mathbf{P}\left(X_n \leq c + \varepsilon\right)\end{aligned}$$

Convergence: Since $X_n \xrightarrow{(d)} X$, we get

$$\lim_{n\to\infty} \mathbf{P}\left(|X_n - c| > \varepsilon\right) = 0$$

Case for which $\xrightarrow{\text{P}}$ yields $\xrightarrow{\text{a.s}}$

Proposition 25.

Consider

- $\{X_n;\ n \geq 1\}$ sequence of random variables

Assume

$$X_n \xrightarrow{\text{P}} X$$

Then there exists a subsequence $\{n_k;\ k \geq 1\}$ such that:

$$X_{n_k} \xrightarrow{\text{a.s}} X$$

Proof of Proposition 25 (1)

Definition of n_k: Recursively we set

$$n_k = \inf\left\{ n > n_{k-1};\ \mathbf{P}\left(|X_n - X| > \frac{1}{k} \right) \le \frac{1}{k^2} \right\}$$

Some notation: For $\varepsilon > 0$ define:

$$\begin{aligned} Y_k &= X_{n_k} \\ A_k(\varepsilon) &= \{ |Y_k - X| > \varepsilon \} \end{aligned}$$

Proof of Proposition 25 (2)

Almost sure convergence: We have

$$\sum_{k=\varepsilon^{-1}}^{\infty} \mathbf{P}\left(A_k(\varepsilon)\right) = \sum_{k=\varepsilon^{-1}}^{\infty} \mathbf{P}\left(|X_{n_k} - X| > \frac{1}{k}\right) \leq \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

Thus

$$Y_k \xrightarrow{\text{a.s}} X$$

Case for which $\xrightarrow{P}$ yields $\xrightarrow{L^r}$

Proposition 26.

Consider

- $\{X_n;\, n \geq 1\}$ sequence of random variables

Assume (bounded convergence)

$X_n \xrightarrow{P} X$, and $|X_n| \leq k$ a.s for all $n \geq 1$ and a given $k > 0$

Then for all $r \geq 1$ we have:

$$X_n \xrightarrow{L^r} X$$

Proof of Proposition 26 (1)

Boundedness of X: For $\delta > 0$, set

$$B_\delta = (|X| \le k + \delta)$$

Then for all $n \ge 1$ we have

$$\begin{aligned} \mathbf{P}(B_\delta) &\ge \mathbf{P}(|X - X_n| \le \delta, \, |X_n| \le k) \\ &\ge \mathbf{P}(|X_n| \le k) - \mathbf{P}(|X - X_n| > \delta) \\ &= 1 - \mathbf{P}(|X - X_n| > \delta) \end{aligned}$$

Taking limits in n, δ we get

$$\mathbf{P}(|X| \le k) = 1$$

Proof of Proposition 26 (2)

Decomposition of $X_n - X$: For $\varepsilon > 0$ and $n \geq 1$ set

$$A_{n,\varepsilon} = \{|X_n - X| > \varepsilon\}$$

Then

$$|X_n - X|^r \leq \varepsilon^r \mathbf{1}_{A_{n,\varepsilon}^c} + (2k)^r \mathbf{1}_{A_{n,\varepsilon}}$$

Taking expectations: We obtain

$$\mathbf{E}\left[|X_n - X|^r\right] \leq \varepsilon^r \mathbf{1}_{A_{n,\varepsilon}^c} + (2k)^r \mathbf{P}\left(A_{n,\varepsilon}\right)$$

Taking limits: With $n \to \infty$ and $\varepsilon \to 0$ we end up with

$$\lim_{n\to\infty} \mathbf{E}\left[|X_n - X|^r\right] = 0$$

Outline

Right inverse (1)

Definition 27.

Let $F : \mathbb{R} \to [0,1]$ continuous cdf
We define the right inverse F^{-1} as

$$F^{-1} : (0,1) \to \mathbb{R}, \quad y \mapsto \inf\{a \in \mathbb{R};\, F(a) \geq y\}$$

Right inverse (2)

Remarks on right inverse:

(i) If F is strictly increasing, F^{-1} is the inverse of F
$\hookrightarrow$ i.e. $F \circ F^{-1} = F^{-1} \circ F = \text{Id}$

(ii) Graphical method to construct F^{-1}:

1. Symmetry wrt diagonal
2. Then erase vertical parts

Example: $F(x) = (x-1)\mathbf{1}_{[1,2)}(x) + \mathbf{1}_{[2,\infty)}(x)$
$\hookrightarrow F^{-1}(y) = (1+y)\mathbf{1}_{(0,1)}(y)$

Right inverse (3)

More remarks:

(iii) Interpretation:

- In above example, $F \equiv$ cdf of $\mathcal{U}([1,2])$
- Domain of interest: $x \in [1,2]$
- In this domain, we do have $F^{-1}(F(x)) = x$

(iv) Generalization:

If $\mu(dx) = f(x)\,dx$ with $\mathrm{Supp}(f) = [a,b]$, then

- F is strictly increasing on $[a,b]$
- $F : (a,b) \to (0,1)$ is invertible
- One can ignore the set $(a,b)^c$ in order to compute F^{-1}

Inverse method for simulation

Proposition 28.

Let

- μ a continuous probability measure on $\mathbb{R}$
- $F(x) = \mu((-\infty, x])$ with right inverse F^{-1}
- $U \sim \mathcal{U}([0,1])$

Then

$X = F^{-1}(U)$ is distributed according to μ

Proof of Proposition 28 (1)

Strategy: We will prove that

$$\begin{aligned} \mathbf{P}(X \leq x) &= \mathbf{P}(F^{-1}(U) \leq x) \\ &\overset{(*)}{=} \mathbf{P}(U \leq F(x)) = F(x) \end{aligned}$$

Details for $(*)$

We wish to show that for $x \in \mathbb{R}$,

$$\left\{u \in (0,1);\ F^{-1}(u) \le x\right\} = \{u \in (0,1);\ u \le F(x)\}$$

Inclusion $\subset$:

$$\begin{aligned} F^{-1}(u) \le x &\Rightarrow \inf\{a;\ F(a) \ge u\} \le x \\ &\Rightarrow \text{There exists } a_1 \le x \text{ such that } F(a_1) \ge u \\ &\Rightarrow F(x) \ge F(a_1) \ge u \end{aligned}$$

Inclusion $\supset$:

$$\begin{aligned} u \le F(x) &\Rightarrow F(x) \ge u \\ &\Rightarrow \inf\{a;\ F(a) \ge u\} \le x \\ &\Rightarrow F^{-1}(u) \le x \end{aligned}$$

Examples

Example 1:

Let $\mu = \mathcal{U}([a,b])$. Then on $[a,b]$

$$F(x) = \frac{x-a}{b-a}, \quad \text{and} \quad F^{-1}(y) = a + (b-a)y$$

One can check that $X = a + (b-a)U \sim \mathcal{U}([a,b])$

Example 2:

Let $\mu = \mathcal{E}(\lambda)$. Then on $\mathbb{R}_+$

$$F(x) = \int_0^x \lambda e^{-\lambda u}\, du = 1 - e^{-\lambda x}, \quad \text{and} \quad F^{-1}(y) = -\frac{\ln(1-y)}{\lambda}$$

One can check that $X = -\frac{\ln(1-U)}{\lambda} \sim \mathcal{E}(\lambda)$

Comments on inverse method

Pros:

- Unique call to `rand`
- Excellent simulation method ... when it works!

Cons:

- Explicit computation of F, F^{-1} not always possible
- Typical example: $\mathcal{N}(0,1)$

Examples of application:

Exponential, Weibull, Cauchy

Outline

Skorohod's representation theorem

Proposition 29.

Consider

- $\{X_n;\, n \geq 1\}$ sequence such that $X_n \xrightarrow{(d)} X$

Then one can construct

1. A probability space $(\Omega, \mathcal{F}, \mathbf{P})$
2. Random variables $Y_n : \Omega \to \mathbb{R}$ satisfying $Y_n \overset{(d)}{=} X_n$
3. $Y : \Omega \to \mathbb{R}$ satisfying $Y \overset{(d)}{=} X$

such that the following holds true:

$$Y_n \xrightarrow{a.s} Y$$

Proof of Proposition 29 (1)

Definition of $(\Omega, \mathcal{F}, \mathbf{P})$: We take

$$\Omega = [0,1], \quad \mathcal{F} = \text{ Borel } \sigma\text{-algebra}, \quad \mathbf{P} = \lambda$$

Definition of Y_n and Y: We take

$$Y_n(\omega) = F_n^{-1}(\omega), \qquad Y(\omega) = F^{-1}(\omega)$$

Distributions of Y_n and Y: According to Proposition 28,

$$Y_n \sim F_n, \qquad Y \sim F$$

Proof of Proposition 29 (2)

Claim 1: If ω is a point of continuity of F^{-1}, we have

$$\lim_{n\to\infty} Y_n(\omega) = \lim_{n\to\infty} F_n^{-1}(\omega) = F^{-1}(\omega) = Y(\omega) \tag{1}$$

Proof of claim 1: Consider

- $\omega \in [0,1]$
- x point of continuity of F such that $Y(\omega) - \varepsilon < x < Y(\omega)$

We have

$$\begin{aligned} F^{-1}(\omega) > x &\implies F(x) < \omega \\ &\implies F_n(x) < \omega, \quad \text{for large } n \\ &\implies x < F_n^{-1}(\omega), \quad \text{for large } n \end{aligned}$$

Proof of Proposition 29 (3)

Proof of claim 1 - ctd: We have seen, for n large enough,

$$Y(\omega) - \varepsilon < x < F_n^{-1}(\omega) \quad \left(\Longrightarrow \quad F_n^{-1}(\omega) > Y(\omega) - \varepsilon\right)$$

Partial conclusion: We get

$$\liminf_{n\to\infty} Y_n(\omega) > Y(\omega) - \varepsilon, \quad \text{for all } \varepsilon > 0$$
$$\Longrightarrow \liminf_{n\to\infty} Y_n(\omega) \geq Y(\omega)$$

Proof of Proposition 29 (4)

Proof of claim 1 - ctd: We have proved

$$\liminf_{n\to\infty} Y_n(\omega) \geq Y(\omega)$$

Along the same lines, for $\omega' > \omega$ one has

$$\limsup_{n\to\infty} Y_n(\omega) \leq Y(\omega')$$

Conclusion: Claim 1 is true, that is
$\hookrightarrow$ If ω is a point of continuity of F^{-1}, we have

$$\lim_{n\to\infty} Y_n(\omega) = Y(\omega) \tag{2}$$

Proof of Proposition 29 (5)

Almost sure convergence: Let

$$D = \left\{\text{points of discontinuity of } F^{-1}\right\}$$

Since F^{-1} non decreasing,

$$\mathbf{P}(D) = \lambda(D) = 0$$

Hence

$$Y_n \xrightarrow{\text{a.s}} Y$$

Characterization of convergence in distribution

Proposition 30.

Consider

- $\{X_n;\, n \geq 1\}$ sequence of random variables

Then the statements 1-2-3 are equivalent:

1. $X_n \xrightarrow{(d)} X$
2. For any $f \in C_b(\mathbb{R})$, we have
$$\lim_{n\to\infty} \mathbf{E}\left[f(X_n)\right] = \mathbf{E}\left[f(X)\right]$$
3. $u \mapsto \mathbf{E}\left[e^{\imath u X}\right]$ cont. at 0, and for every $u \in \mathbb{R}$ we have
$$\lim_{n\to\infty} \mathbf{E}\left[e^{\imath u X_n}\right] = \mathbf{E}\left[e^{\imath u X}\right]$$

Proof of Proposition 30 (1)

Application of Skorohod: One can find

$$Y_n \stackrel{(d)}{=} X_n, \qquad Y \stackrel{(d)}{=} X$$

such that

$$Y_n \xrightarrow{\text{a.s}} Y$$

Convergence of $g(Y_n)$: Since g is continuous, we have

$$g(Y_n) \xrightarrow{\text{a.s}} g(Y)$$

Proof of $1 \Longrightarrow 2$: By bounded convergence,

$$\mathbf{E}[g(Y_n)] \longrightarrow \mathbf{E}[g(Y)]$$

Proof of Proposition 30 (2)

Next step:

Proof of $1 \implies 2$

Approximation of $\mathbf{1}_{(-\infty,x]}$: For $\varepsilon > 0$ we set

$$g_{\varepsilon,x}(y) = \begin{cases} 1, & \text{if } y \leq x \\ 0, & \text{if } y \geq x + \varepsilon \\ \text{linear}, & \text{if } x \leq y \leq x + \varepsilon \end{cases}$$

Upper bound for F_n: For $x \in \mathbb{R}$ we have

$$g_{x,\varepsilon}(y) \geq \mathbf{1}_{(y \leq x)}$$
$$\implies F_n(x) = \mathbf{E}\left[\mathbf{1}_{(X_n \leq x)}\right] \leq \mathbf{E}\left[g_{\varepsilon,x}(X_n)\right]$$

Proof of Proposition 30 (3)

Taking lim sup: Since we assume 2 holds,

$$\begin{aligned}\limsup_{n\to\infty} F_n(x) &\leq \limsup_{n\to\infty} \mathbf{E}\left[g_{x,\varepsilon}(X_n)\right] \\ &\leq \mathbf{P}\left(X \leq x+\varepsilon\right) \\ &= F(x+\varepsilon)\end{aligned}$$

Taking limits in ε: For all x we end up with

$$\limsup_{n\to\infty} F_n(x) \leq F(x)$$

Proof of Proposition 30 (4)

Taking lim inf: By considering $g_{x-\varepsilon,\varepsilon}$ we obtain

$$\begin{aligned}\liminf_{n\to\infty} F_n(x) &\geq \liminf_{n\to\infty} \mathbf{E}\left[g_{x-\varepsilon,\varepsilon}(X_n)\right] \\ &\geq \mathbf{P}\left(X \leq x-\varepsilon\right) \\ &= F(x-\varepsilon)\end{aligned}$$

Taking limits in ε: For a continuity point x of F, we get

$$\liminf_{n\to\infty} F_n(x) \geq F(x)$$

Conclusion: For a continuity point x of F, we have

$$\lim_{n\to\infty} F_n(x) = F(x)$$