

Notes of Math 733: Probability Theory

Instructor: Timo Seppäläinen

YI WEI

Sep 2024

Contents

1	Probability Space	1
2	Laws of Large Numbers	7
2.1	Independence	7
2.1.1	product measures	10
2.1.2	Construction of probability spaces with desired independent r.v.'s	13
2.2	Strong Law Large Number (2.4 in Durett)	14
3		22

1 Probability Space

[Date: Sep 5, 2024]

Setup (Undergraduate level):

Ω sample space: set of all the individual outcomes

\mathcal{F} event space: appropriate collection of subsets of Ω

P : a function on a subsets of Ω , $P(A)$ = the probability of the set (event) A

Axiom 1.1.

$$P\left(\bigcup_k A_k\right) = \sum_k P(A_k) \quad \text{whenever } A_k \text{ is a pairwise disjoint sequence of events}$$

Example 1.1.

1. roll a dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ = power set of Ω = collection of all subset of Ω

2. # of customers to a service station in some fixed time interval

$$\Omega = \mathbb{Z}_{\geq 0}$$

$$\mathcal{F} = \mathcal{P}(\Omega)$$

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k \in \Omega$$

$$P(A) = \sum_{k \in A} e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } A \subseteq \Omega$$

3. Choose uniformly random real number from $[0, 1]$

$$P(x) = 0 \quad \forall x \in [0, 1]$$

if $0 \leq a < b \leq 1$:

$$P([a, b]) = b - a$$

4. Flip a fair coin for infinitely many times, 0 = heads, 1 = tails:

$$\Omega = \{0, 1\}^{\mathbb{Z}_{\geq 0}}$$

$$P\{w : x_1 = a_1, x_2 = a_2, \dots, x_n = a_n\} = 2^{-n} \quad (*)$$

From this: $P\{w\} = 0 \quad \forall w \in \Omega$

Exercise 1.1. how to prove Ω is uncountable: diagonal principle

Definition 1.1. Let X be a space. A σ -algebra on X is a collection \mathcal{A} of subsets of X that satisfies these properties:

1. $\emptyset \in \mathcal{A}$
2. $A \in \mathcal{A} \implies A^C \in \mathcal{A}$
3. $\{A_k\}_{k=1}^{\infty} \implies \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$

And we call (X, \mathcal{A}) is a measurable space.

Definition 1.2. Given (X, \mathcal{A}) A measure is a function $u : \mathcal{A} \rightarrow [0, \infty]$ such that:

1. $P(\emptyset) = 0$
2. $u(\bigcup_k A_k) = \sum_{k=1} u(A_k)$ for a pairwise disjoint sequence $\{A_k\}_k \subseteq \mathcal{A}$

(X, \mathcal{A}, u) is a measure space.

Definition 1.3. If X is a metric space, its Borel σ -algebra \mathcal{B}_X is by definition the smallest σ -algebra containing all the OPEN subsets of X .

Definition 1.4. Lebesgue measure m on \mathbb{R}^d is the measure that satisfies

$$m\left(\prod_{i=1}^d [a_i, b_i]\right) = \prod_{i=1}^d (b_i - a_i)$$

Definition 1.5. A probability space (Ω, \mathcal{F}, P) is a measure space such that $P(\Omega) = 1$.

Example 1.2. Example of product σ -algebra from example 1.1. 4:

\mathcal{F} = product σ -algebra = smallest σ -algebra that contains all sets of the type

$$\{w : x_1 = a_1, \dots, x_n = a_n\} \quad , n \in \mathbb{Z}_{>0}, a_1, \dots, a_n \in \{0, 1\}.$$

P obtained from Eq. *

Definition 1.6. Let (X, \mathcal{A}) , (Y, \mathcal{B}) be measurable space, and $f : X \rightarrow Y$ be a function.

We say f is a measurable function if:

$$f^{-1}(B) = \{x \in X : f(x) \in B\} \subseteq \mathcal{A}, \quad \forall B \in \mathcal{B}$$

A random variable X is a measurable function:

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

Example 1.3. flip of a fair coin $\Omega = \{w = (x_1, x_2) : x_1, x_2 \in \{0, 1\}\}$, 0 = heads, 1 = tails:

$$X_1(w) = x_1 \quad \text{outcome of the first flip}$$

$$X_2(w) = x_2 \quad \text{outcome of the second flip}$$

We define $Y(w) = X_1(w) + X_2(w) = \#$ of tails in the two flips

The information contained in $Y(w)$ is represented by σ -algebra generated by Y defined as follows:

$$\begin{aligned} \sigma(Y) &= \{\{Y \in B\} : B \in \mathcal{B}_{\mathbb{R}}\} \\ &= \left\{ \emptyset, \Omega, \{(0, 0)\}, \{(0, 1), (1, 0)\}, \{(1, 1)\} \text{ and the unions of these sets} \right\} \subsetneq \mathcal{F} \end{aligned}$$

[Date: Sep 10, 2024]

1. push-forward: (X, \mathcal{A}, μ) is a measure space, and (Y, \mathcal{B}) is a measurable space. And there is a $f : X \rightarrow Y$. The push-forward of μ is the measure ν on (Y, \mathcal{B}) defined by $\nu(B) = \mu(f^{-1}(B))$

Exercise 1.2. Check ν is a measure.

2. Absolute continuity: Let μ, λ be measures on (X, \mathcal{A}) . Then μ is absolute continuous w.r.t λ if $\lambda(A) = 0 \implies \mu(A) = 0 \quad \forall A \in \mathcal{A}$.

Remark. $\mu \ll \lambda$. If $\mu \ll \lambda$, then there exists a measurable function $f : X \rightarrow \mathbb{R}_{\geq 0}$ s.t.

$$\mu(A) = \int_A f d\lambda \quad \forall A \in \mathcal{A}$$

This is called Radon-Nikodym derivative $f(x) = \frac{d\mu}{d\lambda}(x)$

Definition 1.7. Let $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be a random variable. The distribution of X is the $\mu = P \circ X^{-1}$, i.e.,

$$\mu(B) = P\{w \in \Omega : X(w) \in B\} \quad \text{for } B \in \mathcal{B}$$

In short: $P\{X \in B\} = P(X \in B)$

Definition 1.8. The CDF of X is the function F on \mathbb{R} defined by

$$F(x) = P(X \leq x) = \mu(-\infty, x]$$

Definition 1.9. If $\mu \ll$ Lebesgue measure, then X has a density function f which satisfies

$$P(a < X \leq b) = \int_a^b f(x) dx = \mu(a, b] = F(b) - F(a)$$

Remark. A discrete random variable has at most countably many values, and since individual pts have positive probability

$$\mu\{k\} = P(X = k) > 0 = \text{leb}\{x\}$$

Then we know $\mu \ll \text{Leb}$ fails and X has no density function.

Definition 1.10. The expectation of a r.v. X is defined by

$$EX = \int_{\Omega} X dP$$

Remark. Abstract Lebesgue integral on (Ω, \mathcal{F}, P)

Definition 1.11. If $A \in \mathcal{F}$ is an event, its indicator random variable is

$$\mathbf{1}_A(w) = \begin{cases} 1, & \text{if } w \in A, \\ 0, & \text{if } w \notin A. \end{cases}$$

We know

$$\begin{aligned} E[\mathbf{1}_A] &= 0 \cdot P\{\mathbf{1}_A = 0\} + 1 \cdot P\{\mathbf{1}_A = 1\} \\ &= P(A) \end{aligned}$$

Example 1.4.

$$X \sim \text{Poisson}(\lambda) \implies E[g(X)] = \sum_{k=0}^{\infty} g(k) \frac{e^{-\lambda} \lambda^k}{k!}$$

$$X \sim \text{Exp}(\lambda) \implies E[g(X)] = \int_0^{\infty} g(x) \lambda e^{-\lambda} dx$$

Theorem 1.2.

Key result:

$$E[f(X)] := \int_{\Omega} f(X) dP = \int_{\mathbb{R}} f d\mu$$

Here: X is a r.v. on (Ω, \mathcal{F}, P) , $\mu = P \circ X^{-1}$ = distribution of X , $f : \mathbb{R} \rightarrow \mathbb{R}$ is a Borel function
 $f(X(w)) = (f \circ X)(w)$

Proof.

$$1. f = \mathbf{1}_B, B \in \mathcal{B}_{\mathbb{R}}.$$

Remark. Notation: $\int_{\Omega} \mathbf{1}_B(X(w)) P(dw)$ (same as $dP(w)$)

$$\begin{aligned} \int_{\Omega} \mathbf{1}_B(X(w)) P(dw) &= \int_{\Omega} \mathbf{1}_{X^{-1}(B)}(w) dx \\ &= P(X^{-1}(B)) = \mu(B) = \int_{\mathbb{R}} \mathbf{1}_B d\mu \end{aligned}$$

$$2. f = \sum_{i=1}^n a_i \mathbf{1}_{B_i}, a_1, \dots, a_n \in \mathbb{R}, B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$$

$$\begin{aligned} \int_{\Omega} \sum_{i=1}^n a_i \mathbf{1}_{B_i}(X) dP &= \sum_{i=1}^n a_i \int_{\Omega} \mathbf{1}_{B_i}(X) dP \\ &= \sum_{i=1}^n a_i \int_{\mathbb{R}} \mathbf{1}_{B_i} d\mu \\ &= \int_{\mathbb{R}} \sum_{i=1}^n a_i \mathbf{1}_{B_i} d\mu \end{aligned}$$

$$3. f \geq 0, \exists \text{ simple function } 0 \leq f_n$$

$$\begin{aligned} \int_{\Omega} f(X) dP &= \lim_{n \rightarrow \infty} \int_{\Omega} f_n(X) dP \quad (M.C.T.) \\ &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n d\mu \\ &= \int_{\mathbb{R}} f d\mu \end{aligned}$$

Remark. $f_n(x) = \sum_{k=0}^{n(2^n-1)} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\}} + n \mathbf{1}_{\{f(x) \geq n\}}$

$$4. \text{ For general } f : \mathbb{R} \rightarrow \mathbb{R} = f^+ - f^- \text{ Borel function where } f^+, f^- \geq 0$$

$$\int_{\Omega} f(X) dP = \int_{\Omega} f^+(X) dP - \int_{\Omega} f^-(X) dP$$

$$\begin{aligned}
&= \int_{\mathbb{R}} f^+ d\mu - \int_{\mathbb{R}} f^- d\mu \\
&= \int_{\mathbb{R}} f d\mu
\end{aligned}$$

□

Example 1.5. 1. $X \sim \text{Poisson}(\lambda)$, μ = distribution of X . We know $\mu(B) = \sum_{k:k \in B} e^{-\lambda} \frac{\lambda^k}{k!} \implies \mu(\mathbb{R} \setminus \mathbb{Z}_{\geq 0}) = 0$. Then we have:

$$\begin{aligned}
E[e^{-tX}] &= \int_{\mathbb{R}} e^{-tx} \mu(dx) \\
&= \int_{\mathbb{Z}_{\geq 0}} e^{-tx} \mu(dx) \\
&= \sum_{k \in \mathbb{Z}_{\geq 0}} \int_{\{k\}} e^{-tx} \mu(dx) \\
&= \sum_{k \geq 0} e^{-tk} e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} e^{\lambda e^{-t}} \\
&= e^{\lambda(e^{-t} - 1)}
\end{aligned}$$

2. $X \sim \text{Exp}(\lambda)$

$$\begin{aligned}
E[e^{-tX}] &= \int_{\mathbb{R}} e^{-tx} \mu(dx) = \int_{[0, \infty)} e^{-tx} \lambda e^{-\lambda x} dx \\
&= \lim_{M \rightarrow \infty} \int_{[0, M]} \lambda e^{-(t+\lambda)x} dx = \lim_{M \rightarrow \infty} \lambda \int_0^M e^{-(t+\lambda)x} dx \\
&= \lim_{M \rightarrow \infty} \left(-\frac{\lambda}{t+\lambda} e^{-(t+\lambda)x} \right) \Big|_0^M \\
&= \lim_{M \rightarrow \infty} \left(\left(-\frac{\lambda}{t+\lambda} \right) e^{-(t+\lambda)M} + \frac{\lambda}{t+\lambda} \right) \\
&= \frac{\lambda}{t+\lambda}
\end{aligned}$$

2 Laws of Large Numbers

[Date: Sep 12, 2024]

2.1 Independence

Perhaps you recall this: events A and B are independent if $P(A \cap B) = P(A)P(B)$

Definition 2.1. Let Ω, \mathcal{F}, P be a probability space.

Let $\mathcal{F}_1, \dots, \mathcal{F}_n$ be sub- σ -algebra of \mathcal{F} . (Means each \mathcal{F}_i is a σ -algebra and $\mathcal{F}_i \subseteq \mathcal{F}$)

Then we say $\mathcal{F}_1, \dots, \mathcal{F}_n$ are independent if $\forall A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$, then

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

Now r.v.'s X_1, \dots, X_n on Ω are independent if the σ -algebra $\sigma(X_1), \dots, \sigma(X_n)$ are independent. Equivalently, \forall measurable sets in the range space,

$$P\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n P\{X_i \in B_i\}$$

Events A_1, \dots, A_n are independent if the r.v.'s $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ are independent.

And arbitrary collection $\{\mathcal{F}_\beta : \beta \in \mathcal{J}\}$ of sub- σ -algebra is independent if \forall distinct $\beta_1, \dots, \beta_n \in \mathcal{J}$, $\mathcal{F}_{\beta_1}, \dots, \mathcal{F}_{\beta_n}$ are independent.

Claim 2.1. Fact: X_1, \dots, X_n are independent, then so are $f_1(X_1), \dots, f_n(X_n)$

Remark. Why product?

X, Y discrete r.v.'s. We're interested in the event $\{X = k\}$. Suppose we learn that $Y = m$. We replace P with $P(\cdot, Y = m)$ defined by $P(A|Y = m) = \frac{P(A \cap \{Y=m\})}{P(Y=m)}$

When is $P(X = k) = P(X = k|Y = m)$?

$$\begin{aligned} P(X = k) &= P(X = k|Y = m) \\ \iff P(X = k)P(Y = m) &= P(X = k, Y = m) \end{aligned}$$

We need some notions/tool to check easily if two r.v.'s are independent.

1. Develop a simpler criterion for checking independence of a given collection of r.v.'s.
2. To construct a probability space with desired independent r.v.'s.

Example 2.1. Let X_1, X_2, X_3 be independent *Bernolli*(p) r.v.'s.

$$P(X_i = 1) = p = 1 - P(X_i = 0)$$

Consider the following events:

$$\begin{cases} \{X_1 + X_2 = 1\} \\ \{X_2 + X_3 = 1\} \end{cases}$$

Firstly we have:

$$P(X_1 + X_2 = 1) = P(01) + P(10) = 2p(1 - p) = P(X_2 + X_3 = 1)$$

And we have:

$$P(X_1 + X_2 = 1, X_2 + X_3 = 1) = P(101) + P(010) = p^2(1 - p) + p(1 - p)^2 = p(1 - p)$$

If the two events are independent, we have:

$$\begin{aligned} P(X_1 + X_2 = 1, X_2 + X_3 = 1) &= P(X_1 + X_2 = 1) \cdot P(X_2 + X_3 = 1) \\ \iff p(1 - p) &= 4p^2(1 - p)^2 \\ \iff p(1 - p) &= \frac{1}{4}, \quad p = 0, \text{ or } p = 1 \\ \iff p &= \frac{1}{2}, 0, \text{ or } 1 \end{aligned}$$

Theorem 2.2. Let $\mathcal{A}_1, \dots, \mathcal{A}_n$ be subcollection of \mathcal{F} . Assume that each \mathcal{A}_i is closed under intersection, which means $(A, B \in \mathcal{A}_i \implies A \cap B \in \mathcal{A}_i)$ and $\Omega \in \mathcal{A}_i$. Assume that the probability $P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) \quad \forall A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$. Then the σ -algebra $\sigma(\mathcal{A}_1), \dots, \sigma(\mathcal{A}_n)$ are independent.

Example 2.2. Collection of sets which can generate Borel-algebra:

$A_i = \{(a, b) : -\infty < a < b < \infty\}$, then $\sigma(A_i) = \mathcal{B}_{\mathbb{R}}$.

Or you can take $(-\infty, b]$

The tool for proving the theorem: Dynkin's $\pi - \lambda$ theorem.

Definition 2.2. Let \mathcal{A} be a collection of subset of Ω

1. \mathcal{A} is a π -system if it is closed under intersections.
2. \mathcal{A} is a λ -system if it has the following three properties:

- (a) $\Omega \in \mathcal{A}$
- (b) $\forall A, B \in \mathcal{A}$ and $A \subseteq B \implies B \setminus A \in \mathcal{A}$
- (c) If $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq \dots$ and each $A_i \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Theorem 2.3. Suppose \mathcal{P} is a π -system, \mathcal{L} is a λ -system and $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$

We use theorem 2.3 to prove theorem 2.2.

Proof of theorem 2.2:

Fix $A_2 \in \mathcal{A}_2, \dots, A_n \in \mathcal{A}_n$, set $\mathcal{F} = A_2 \cap \dots \cap A_n$

$$\mathcal{L} = \{A \in \mathcal{F} : P(A \cap \mathcal{F}) = P(A)P(\mathcal{F})\}$$

Claim 2.4. $\mathcal{A}_1 \subseteq \mathcal{L}$.

Proof of Claim 2.4.

Check that $P(\mathcal{F}) = \prod_{i=2}^n P(A_i)$

Take $A_1 = \Omega$

Let $A_1 \in \mathcal{A}_1$. $P(A_1 \cap \mathcal{F}) = P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) = P(A_1)P(\mathcal{F})$ □

Claim 2.5. \mathcal{L} is a λ -system.

Proof of Claim 2.5.

1. $\Omega \in \mathcal{A}_1 \subseteq \mathcal{L}$
2. Let $A, B \in \mathcal{L}, A \subseteq B$. We want $B \setminus A \in \mathcal{L}$.

$$P\left((B \setminus A) \cap \mathcal{F}\right) = P\left((B \cap \mathcal{F}) \setminus (A \cap \mathcal{F})\right) = P(B \cap \mathcal{F}) - P(A \cap \mathcal{F})$$

3. Let $\mathcal{L} \ni A_i \nearrow A$. We want: $A \in \mathcal{L}$

$$P(A \cap \mathcal{F}) = \lim_{n \rightarrow \infty} P(A_n \cap \mathcal{F}) \quad \text{because } A_n \cap \mathcal{F} \nearrow A \cap \mathcal{F}$$

We've checked that \mathcal{L} is a λ -system. So $\sigma(\mathcal{A}_1) \subseteq \mathcal{L}$ □

We continue the proof of theorem 2.2:

Then $P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) \quad \forall A_1 \in \sigma(\mathcal{A}_1), A_2 \in \mathcal{A}_2, \dots, A_n \in \mathcal{A}_n$

We can use the same argument to upgrade each \mathcal{A}_i in turn to $\sigma(\mathcal{A}_i)$. At the end we have the product properties for all members of $\sigma(\mathcal{A}_1), \dots, \sigma(\mathcal{A}_n)$ □

Corollary 2.6. \mathbb{R} -valued r.v.'s X_1, \dots, X_n are independent iff

$$P\left(\bigcap_{i=1}^n \{X_i \leq s_i\}\right) = \prod_{i=1}^n P\{X_i \leq s_i\}$$

[Date: Sep 17, 2024]

Today:

$$\text{Independent r.v's} \begin{cases} \text{product measure} \\ \text{convolutions} \end{cases}$$

2.1.1 product measures

Definition 2.3. Suppose $(X_1, \mathcal{A}_1, \mu_1), \dots, (X_n, \mathcal{A}_n, \mu_n)$ are σ -finite measure spaces. The product measure space (X, \mathcal{A}, μ) is defined as follows:

$$X = \prod_{i=1}^n X_i = \text{the Cartesian product, } \mathcal{A} = \text{product } \sigma\text{-algebra} = \otimes_{i=1}^n \mathcal{A}_i = \sigma\{A_1 \times \dots \times A_n : A_i \in \mathcal{A}_i\}$$

$\mu = \text{product measure} = \otimes_{i=1}^n \mu_i =$ by def the unique measure μ on \mathcal{A} such that

$$\mu(A_1 \times \dots \times A_n) = \prod_{i=1}^n \mu_i(A_i) \quad \forall A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$$

Theorem 2.7 (Tonelli-Fubini Theorem).

$$(n = 2): \int_{X \times Y} f(x, y) \mu \otimes \nu(dx, dy) = \int_Y \left[\int_X f(x, y) \mu(dx) \right] \nu(dy)$$

Suppose each X_i is a metric space w.r.t \mathcal{B}_{X_i} ; also X is a metric space w.r.t \mathcal{B}_X . Relationship of $\otimes_{i=1}^n \mathcal{B}_{X_i}$ & \mathcal{B}_X .

Take $n = 2$: $\mathcal{B}_X \otimes \mathcal{B}_Y = \sigma\{A \times B : A \in \mathcal{B}_X, B \in \mathcal{B}_Y\} = \sigma\{ \underbrace{A \times B}_{\text{open set in } X \times Y} : A \subseteq X \text{ open, } B \subseteq Y \text{ open} \}$

Definition 2.4. Separable metric space has a countable dense subset.

Example 2.3.

1. \mathbb{R}^d
2. $C[0, 1]$
3. $C([0, \infty])$: here the metric $d(f, g) = \sup_{0 \leq x < \infty} |f(x) - g(x)|$ makes not separable!
But $d(f, g) = \sum_{n=1}^{\infty} 2^{-n} (\sup_{0 \leq x \leq n} |f(x) - g(x)| \wedge 1)$

Theorem 2.8 (Proposition 1.5 in Folland).

Fact: If X, Y are separable metric spaces, then $\mathcal{B}_X \otimes \mathcal{B}_Y = \mathcal{B}_{X \times Y}$

Remark. reference: Richard M. Dudley: Real Analysis and Probability, Prop 4.1.7

Definition 2.5. Suppose X_1, \dots, X_n are r.v.'s on (Ω, \mathcal{F}, P) . Let $\mu_i(B) = P(X_i \in B)$, $B \in \mathcal{B}_{\mathbb{R}}$ be the distribution (marginal distribution of X_i) of X_i

$X = (X_1, \dots, X_n)$ is an \mathbb{R}^n -valued random variable and its distribution (joint distribution of X_1, \dots, X_n) is a probability measure μ on \mathbb{R}^n .

Theorem 2.9. X_1, \dots, X_n independent $\iff \mu = \otimes_{i=1}^n \mu_i$

Proof.

1. \implies : Let $A_1 \times \dots \times A_n \in \mathcal{B}_{\mathbb{R}^n}$.

$$\begin{aligned} \mu(A_1 \times \dots \times A_n) &= P\{(X_1, \dots, X_n) \in A_1 \times \dots \times A_n\} \\ &= P\{X_1 \in A_1, \dots, X_n \in A_n\} = \prod_{i=1}^n P(X_i \in A_i) = \prod_{i=1}^n \mu_i(A_i). \\ &\quad \pi - \lambda \text{ thm} \implies \mu = \otimes_{i=1}^n \mu_i \end{aligned}$$

2. \implies Similar

□

Corollary 2.10. If $E|f_i(X_i)| < \infty$ for $i = 1, \dots, n$, X_1, \dots, X_n independent, then

$$E\left[\prod_{i=1}^n f_i(x_i)\right] = \prod_{i=1}^n E[f_i(X_i)]$$

Proof. Note that when X_1, \dots, X_n are independent, then $f(X_1), \dots, f(X_n)$ are independent.

Take $n = 2$. Let $\mu_i = P \circ X_i^{-1}$

$$\begin{aligned} E[f_1(X_1)f_2(X_2)] &= \int_{\mathbb{R}^2} f_1(x_1)f_2(x_2)(\mu_1 \otimes \mu_2)(dx_1 dx_2) \\ &= \int_{\mathbb{R}} \mu_2(dx_2) \int_{\mathbb{R}} \mu_1(dx_1) f_1(x_1) f_2(x_2) \\ &= \int_{\mathbb{R}} \mu_2(dx_2) f_2(x_2) \int_{\mathbb{R}} \mu_1(dx_1) f_1(x_1) \\ &= E[f_1(X_1)] E[f_2(X_2)] \end{aligned}$$

□

Remark. It's OK to mix notation: if $X \perp\!\!\!\perp Y$, then

$$\begin{aligned} E[g(X, Y)] &= \int g d\mu \otimes \nu = \int \nu(dy) \int \mu(dx) g(x, y) \\ &= \int \nu(dy) \mathbb{E}[g(X, y)] \end{aligned}$$

Corollary 2.11. Let $X = (X_1, \dots, X_n)$ have PDF f on \mathbb{R}^n , and let f_i be PDF of X_i for $i = 1, \dots, n$. Then

$$X_1, \dots, X_n \text{ are independent} \iff f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) \quad \text{for Lebesgue almost every } (x_1, \dots, x_n) \in \mathbb{R}^n$$

Definition 2.6 (convolutions). Let μ, ν be Borel probability measure on \mathbb{R} . Their convolution is

$$\mu * \nu(B) = \int_{\mathbb{R}} \mu(B - x) \nu(dx), \quad B \in \mathcal{B}_{\mathbb{R}}$$

Why is $\mu(B - x)$ is measurable?

$$\mu(B - x) = \int_{\mathbb{R}} \mathbf{1}_{B-x}(y) \mu(dy) = \int_{\mathbb{R}} \underbrace{\mathbf{1}_B(x + y)}_{\text{jointly measurable function}}(x, y) \mu(dy)$$

Fubini \implies the interpretation over y leaves a measurable function of the variable x .

We consider the probability meaning of $\mu * \nu$:

$$\begin{aligned} \mu * \nu(B) &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbf{1}_B(x + y) \mu(dx) \right] \nu(dy) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_B(x + y) (\mu \otimes \nu)(dx, dy) \\ &= \mathbb{E}[\mathbf{1}_B(X + Y)] \\ &= P(X + Y \in B) \end{aligned}$$

Let $X \perp Y$, $X \sim \mu, Y \sim \nu$. Then we have $(X, Y) \sim \mu \otimes \nu$

Theorem 2.12. $X \perp Y, X \sim \mu, Y \sim \nu \implies X + Y \sim \mu \times \nu$

What happened

Suppose μ has PDF f , ν has PDF g . Find

$$\begin{aligned} \mu * \nu(A) &= \int_{\mathbb{R}^2} \mathbf{1}_A(x + y) \mu(dx) \nu(dy) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_A(x + y) f(x) g(y) dx dy \\ &= \int_{\mathbb{R}} dx f(x) \int_{\mathbb{R}} \mathbf{1}_A(x + y) g(y) dy \\ &= \int_{\mathbb{R}} dx f(x) \int_{\mathbb{R}} \mathbf{1}_A(y) (y - x) dy \\ &= \int_{\mathbb{R}} dy \mathbf{1}_A(y) \int_{\mathbb{R}} dx f(x) g(y - x) \end{aligned}$$

By definition $f * g(y)$ we see that is the PDF of $\mu * \nu$

Example 2.4. Gaussian density: $f(x_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}$ and $f(x_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-m_2)^2}{2\sigma_2^2}}$.

We have

$$(f_1 * f_2)(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2 + \sigma_2^2)}}$$

[Date: Sep 19, 2024]

2.1.2 Construction of probability spaces with desired independent r.v.'s

2.1.4 section in Durrett

Finite case: Given μ_1, \dots, μ_n Borel probability measure on \mathbb{R} .

Want: independent r.v.'s X_1, \dots, X_n with $X_i \sim \mu_i$

Take $\Omega = \mathbb{R}^n = \{\omega = (x_1, \dots, x_n) : x_i \in \mathbb{R}\}$. $\mathcal{F} = \mathcal{B}_{\mathbb{R}} = \mathcal{B}_{\mathbb{R}}^{\otimes n}$ (X_i has probability distribution μ_i), $P = \otimes_{i=1}^n \mu_i$, $X_i(\omega) = x_i$ ("coordinate r.v.'s coordinate projections").

Given $B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$

$$\begin{aligned} P(X_1 \in B_1, \dots, X_n \in B_n) &= P\{\omega \in \Omega : X_i(\omega) \in B_1, \dots, X_n(\omega) \in B_n\} \\ &= (\otimes_{i=1}^n \mu_i)\{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 \in B_1, \dots, x_n \in B_n\} \\ &= (\otimes_{i=1}^n \mu_i)\left(\prod_{i=1}^n B_i\right) \\ &= \prod_{i=1}^n \mu_i(B_i) \\ &= \prod_{i=1}^n P(X_i \in B_i) \quad \text{by 1} \end{aligned}$$

Intermediate step: pick j , take $B_i = \mathbb{R}$ for $i \neq j$, substitute with the calculation:

$$\begin{aligned} P(X_j \in B_j) &= \prod_{i=1}^n \mu_i(B_i) = \mu_j(B_j) \\ &\implies X_j \sim \mu_j \end{aligned} \tag{1}$$

This all works if we replace $\mathbb{R}, \mathcal{B}_{\mathbb{R}}$ with arbitrary measurable spaces (S_i, \mathcal{A}_i) . The choice of (Ω, \mathcal{F}, P) is not unique at all!

Definition 2.7 (Infinite case). A stochastic process is an dexted collection $\{X_\alpha : \alpha \in \mathcal{J}\}$ of r.v.'s all defined on the same (Ω, \mathcal{F}, P) .

Theorem 2.13 (Kolmogorov's Extension Theorem). (for index set $\mathbb{Z}_{\geq 0}$) Assume that $\forall n \geq 1$, we have a probability measure \mathbf{u}_n on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ and these measures are consistent: $\forall B \in \mathcal{B}_{\mathbb{R}^n} : \mathbf{u}_{n+1}(B) = \mathbf{u}_n(B)$

Let $\Omega = \mathbb{R}^{\mathbb{Z}_{\geq 0}} = \{\omega = (x_i)_{i=1}^\infty : \text{each } x_i \in \mathbb{R}\}$, $\mathcal{F} = \text{product } \sigma\text{-algebra} = \sigma\{A_1 \times \dots \times A_n \times \mathbb{R} \times \mathbb{R} \times \dots : n \in \mathbb{Z}_{>0}, A_1, \dots, A_n \in \mathcal{B}_{\mathbb{R}}\} = \sigma\text{-algebra generated by the projection mapping } X_i(\omega) = x_i, i \in \mathbb{Z}_{>0}, = \text{smallest } \sigma\text{-algebra on } \Omega \text{ under which each } X_i : \Omega \rightarrow \mathbb{R} \text{ is measurable.}$

Then \exists unique probability measure P on Ω such that $P\{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in B\} = \mathbf{u}_n(B) \quad \forall n \in \mathbb{Z}_{>0}, B \in \mathcal{B}_{\mathbb{R}^n}$

Theorem 2.14 (Kolmogorov Extension theorem process version). Given consistent finite-dim distribution $\{\mathbf{u}_n\}_{n \geq 1}$ on $\mathbb{R}^n \quad \forall n, \exists$ a stochastic process $(X_k)_{k \in \mathbb{Z}_{>0}}$ with marginal $(X_1, \dots, X_n) \sim \mathbf{u}_n$

Proof. Take the coordinate process from the previous theorem. □

Generalizations:

1. Instead of \mathbb{R} , we can take any Borel subsets of complete separable metric spaces..
"Polish spaces"
2. The index set can be totally arbitrary. [cf. Dudley's book]

To produce a process $(X_k)_{k \in \mathbb{Z}_{>0}}$ of independent r.v.'s with $X_k \sim \mu_k$, take $\mathbf{u}_n = \mu_1 \otimes \cdots \otimes \mu_n$ in K's extension theorem.

Definition 2.8. An IID process is a process of independent identically distributed r.v.'s.

2.2 Strong Law Large Number (2.4 in Durrett)

Two big goals for IID process $\{X_k\}_{k \in \mathbb{Z}_{>0}}$

Theorem 2.15. If $\mathbb{E}|X_1| < \infty$, then $S_n = X_1 + \cdots + X_n$ satisfies

$$\frac{S_n}{n} \longrightarrow \mathbb{E}X_1 \quad \text{w.p.1}$$

Theorem 2.16. Central Limit Theorem: if $\sigma^2 = \text{Var}(X_1) < \infty$, then

$$P\left\{\frac{S_n - n\mathbb{E}X_1}{\sigma\sqrt{n}} \leq s\right\} \xrightarrow{n \rightarrow \infty} \int_{-\infty}^s \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

Definition 2.9. Let $\{A_n\}$ be a sequence of events in (Ω, \mathcal{F}, P) .

$$\begin{aligned} \{A_n \text{ i.o. (infinitely often)}\} &= \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\} \\ &= \{\omega \in \Omega : \forall m \geq 1, \exists n \geq m \text{ s.t. } \omega \in A_n\} \\ &= \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n (= \limsup A_n) \end{aligned}$$

Theorem 2.17 (1st Borel-Cantelli Lemmas).

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies P(A_n \text{ i.o.}) = 0$$

(1.) Let $N(\omega) = \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = \#$ of events that occur.

$$\begin{aligned} \mathbb{E}[N] &\stackrel{\text{MCT}}{=} \sum_{n=1}^{\infty} \mathbb{E}[\mathbf{1}_{A_n}] = \sum_{n=1}^{\infty} P(A_n) < \infty \\ &\implies P(N = \infty) = 0 \end{aligned}$$

□

(2.)

$$\begin{aligned} P\left(\bigcap_{m \geq 1} \bigcup_{n \geq m} A_n\right) &= \lim_{m \rightarrow \infty} P\left(\bigcup_{n \geq m} A_n\right) \\ &\leq \lim_{n \rightarrow \infty} \sum_{n \geq m} P(A_n) = 0 \quad \text{by convergent series tails} \end{aligned}$$

□

Definition 2.10. (Suppose all defined on the same probability space) $X_n \xrightarrow{a.s.} X$ if $P\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1$

Lemma 2.18. Suppose $\forall \epsilon > 0, \sum_{n=1}^{\infty} P(|x_n - x| \geq \epsilon) < \infty$.

Then $X_n \rightarrow X \quad a.s..$

Proof. Pick any sequences $0 < \epsilon_j \searrow$

$$\begin{aligned} \text{B-C} &\implies P\left(\bigcap_{m \geq 1} \bigcup_{n \geq m} \{|X_n - X| \geq \epsilon_j\}\right) = 0 \\ &\implies 1 = P\left(\bigcup_{m \geq 1} \bigcap_{n \geq m} \{|X_n - X| < \epsilon_j\}\right) \\ &= P\{\exists m < \infty \text{ s.t. } n \geq m \implies |X_n - X| < \epsilon_j\} \\ 1 &= P\left(\bigcap_{j=1}^{\infty} \bigcup_{m \geq 1} \bigcap_{n \geq m} \{|X_n - X| < \epsilon_j\}\right) \\ &= P\{\forall j \exists m, m, n \geq m \implies |X_n - X| < \epsilon_j\} \\ &= P\{X_n \rightarrow X\} \end{aligned}$$

□

Example 2.5. Suppose $\mathbb{E}|Y_n| \leq 2^{-n}$. Then $Y_n \xrightarrow{a.s.} 0$.

Proof. $\sum_{n \geq 1} P(|Y_n| \geq \epsilon) \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}|Y_n|}{\epsilon} \leq \frac{1}{\epsilon} \sum_{n=1}^{\infty} 2^{-n} < \infty$

□

Remark (Markov-Chebyshev). Suppose r.v. $Z \geq 0, a > 0$:

$$P(Z \geq a) = \mathbb{E}[\mathbf{1}_{Z \geq a}] \leq \mathbb{E}\left[\frac{Z}{a} \mathbf{1}_{Z \geq a}\right] \stackrel{Z \geq 0}{\leq} \mathbb{E}\left[\frac{Z}{a}\right]$$

Lemma 2.19 (Borel-Cantelli). $\{A_n\}_{n \geq 1}$ a sequence of events on (Ω, \mathcal{F}, P) and $\sum_n P(A_n) < \infty \implies P\left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n\right) = 0$

Lemma 2.20 (2nd Borel-Cantelli). Let $\{A_n\}_{n \geq 1}$ be independent events. Then

$$\sum_n P(A_n) < \infty \implies P\left(\bigcap_{n=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\right) = 1$$

Proof. Let $M < N$.

$$\begin{aligned}
P\left(\bigcap_{n=M}^N A_n^c\right) &= \prod_{n=M}^N (1 - P(A_n)) \leq \prod_{n=M}^N e^{-P(A_n)} \\
&= e^{-\sum_{n=M}^N P(A_n)} \xrightarrow{N \rightarrow \infty} 0 \\
&\implies P\left(\bigcap_{n=M}^{\infty} A_n^c\right) = 0 \\
&\implies 1 = P\left(\bigcup_{n=M}^{\infty} A_n\right) = 1 \quad \text{true } \forall M
\end{aligned}$$

So

$$P\left(\bigcap_{M=1}^{\infty} \bigcup_{n=M}^{\infty} A_n\right) = 1 \quad (2)$$

□

Example 2.6. Roll a fair dice ∞ often. X_n = outcome of n -th roll.

$$\begin{aligned}
\sum_{n=1}^{\infty} P(X_n = 6) &= \sum_{n=1}^{\infty} \frac{1}{6} = \infty \\
&\xrightarrow{2\text{nd} B-C} P(X_n = 6 \text{ i.o.}) = 1
\end{aligned}$$

Definition 2.11. $X_n \xrightarrow{a.s.} X$ if $P\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1$

Definition 2.12. X_n converges to X in probability ($X_n \xrightarrow{P} X$) if $\forall \epsilon > 0$

$$P(|X_n - X| \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0$$

Remark (Observation). $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$

$$\begin{aligned}
1 &= P(X_n \rightarrow X) = P\left(\bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \left\{|X_n - X| \leq \frac{1}{k}\right\}\right) \\
&\leq P\left(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \left\{|X_n - X| \leq \frac{1}{k}\right\}\right) \quad \forall k \in \mathbb{Z}_{>0} \quad (\text{Notice that this increases with } m) \\
&= \lim_{m \rightarrow \infty} P\left(\bigcap_{n=m}^{\infty} \left\{|X_n - X| \leq \frac{1}{k}\right\}\right) \\
&\leq \underbrace{\lim_{m \rightarrow \infty} P(|X_m - X| \leq \frac{1}{k})}_{\text{this limit} = 1 \quad \forall k}
\end{aligned}$$

Theorem 2.21.

$X_n \xrightarrow{P} X \implies$ Every subsequence $\{X_{n(m)}\}_{m=1}^{\infty}$ has a further subsequence $X_{n(m_k)} \xrightarrow[k \rightarrow \infty]{a.s.} X$

Proof. • (\implies) Assumption is : $\forall \epsilon > 0, P(|X_n - X| \geq \epsilon) \longrightarrow 0$

Let subseq $\{n(m)\}_{m=1}^\infty$ be given. Pick $n(m_1) < n(m_2) < \dots < n(m_k) < \dots$ such that

$$\begin{aligned} P(|X_{n(m_k)} - X| \geq 2^{-k}) &< 2^{-k} \\ \sum_{k=1}^{\infty} 2^{-k} &< \infty \implies \text{by B-C, w.p.1.} \\ \exists k_0 = k_0(\omega) \text{ s.t. } k \geq \underbrace{k_0(\omega)}_{\text{random index}} &\implies |X_{n(m_k)} - X| < 2^{-k} \quad \omega \in \Omega \end{aligned}$$

Remark. k_0 is random means that it is a function of ω .

- (\impliedby) pf by contradiction: suppose $X_n \xrightarrow{P} X$ fails. Then $\exists \epsilon > 0$ s.t. $P(|X_n - X| \geq \epsilon) \not\longrightarrow 0$. Fails. So \exists subseq $n(m)$ and $\delta > 0$ s.t. $P(|X_{n(m)} - X| \geq \epsilon) \geq \delta$. But by assumption \exists further subsequence converges a.s. $X_{n(m_k)} \xrightarrow{a.s.} X \implies X_{n(m_k)} \xrightarrow{P} X \implies P(|X_{n(m_k)} - X| \geq \epsilon) \longrightarrow 0$ \square

Theorem 2.22 (SLLN and WLLN). SLLN:

$$P\left(\frac{S_n}{n} \longrightarrow \mathbb{E}X_1\right) = 1$$

where $S_n = \sum_{i=1}^n X_i$ IID $\{X_k\}$ WLLN: $\frac{S_n}{n} \xrightarrow{P} \mathbb{E}X_1$

Here we can write $\frac{S_n - n\mathbb{E}X_1}{n} = \frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{a.s.} 0$

Lemma 2.23 (Chebyshev's Inequality).

$$\begin{aligned} P(|X - \mathbb{E}X| \geq a) &= P(|X - \mathbb{E}X|^2 \geq a^2) \\ &\leq \frac{\mathbb{E}[|X - \mathbb{E}X|^2]}{a^2} \\ &= \frac{\text{Var}(X)}{a^2} \end{aligned}$$

Remark. Why do we wanna use the square? Because L^2 is the Hilbert space but L^1 is just a Banach space.

Definition 2.13. X and Y are uncorrelated is $0 = \text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$.

We know that $X \perp Y \implies X$ and Y uncorrelated.

Computation:

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}X + \mathbb{E}Y)^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - (\mathbb{E}X)^2 - 2\mathbb{E}X\mathbb{E}Y - (\mathbb{E}Y)^2 = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

Theorem 2.24 (SLLN). Let $\{X_k\}_{k=1}^{\infty}$ be pairwise independent, identically distributed, integrable. Let $S_n = \sum_{k=1}^n X_k$. Then $P\{\frac{S_n}{n} \longrightarrow EX_1\} = 1$

Proof.

Truncation: define $Y_k = X_k \cdot \mathbf{1}_{\{|X_k| \leq k\}}$. Let $T_n = \sum_{k=1}^n Y_k$

Lemma 2.25 ((a)). The theorem will follow from showing that

$$\frac{T_n}{n} \xrightarrow{a.s.} EX_1$$

proof of lemma (a.)

$$\begin{aligned} \sum_{k=1}^{\infty} P(X_k \neq Y_k) &= \sum_{k=1}^{\infty} P(|X_k| > k) \\ &= \sum_{k=1}^{\infty} \int_{k-1}^k P(|X_k| > k) dt \\ &\leq \sum_{k=1}^{\infty} \int_{k-1}^k P(|X_1| > t) dt \\ &= \int_0^{\infty} P(|X_1| > t) dt \\ &= E|X_1| < \infty \end{aligned}$$

Notice for $\int_0^{\infty} P(|X_1| > t) dt$:

$$\begin{aligned} \int_0^{\infty} E\mathbf{1}_{|X_1| > t} dt &= E \int_0^{\infty} \mathbf{1}_{|X_1| > t} dt \\ &= E \int_0^{|X_1|} dt = E|X_1| \end{aligned}$$

B-C $\implies \exists$ random k_0 s.t.

$$k > k_0 \implies X_k = Y_k$$

Suppose $\frac{T_n}{n} \longrightarrow EX_1$ a.s.

$$\begin{aligned} \frac{S_n}{n} &= \frac{T_n}{n} + \frac{S_n - T_n}{n} \\ \implies \frac{|S_n - T_n|}{n} &\leq \frac{1}{n} \sum_{k=1}^n |X_k - Y_k| \\ &\leq \frac{1}{n} \sum_{k=1}^{k_0(\omega)} |X_k(\omega) - Y_k(\omega)| \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

[Date: Sep 26, 2024] Last time we observed $EX = \int_0^{\infty} P(X > s) ds$ for $X \geq 0$.

This generalizes: let $X \geq 0$, $h : [0, \infty) \rightarrow \mathbb{R}$ nondecreasing, $h(0) = 0$, $h \in C^1$:

$$E[h(x)] = E[h(x) - h(0)] = E\left[\int_0^X h'(x)dx\right] = E\left[\int_0^\infty h'(x)\mathbf{1}_{X>0}dx\right] \quad (3)$$

$$= \int_0^\infty h'(x)P(X > 0)dx \quad (4)$$

□

Lemma 2.26 (b.). $\sum_{k=1}^\infty \frac{\text{Var}(Y_k)}{k^2} \leq 4E|X_1| < \infty$

Proof.

$$\text{Var}(Y_k) \leq E(Y_k^2) = \int_0^\infty 2y \underbrace{P(|Y_k| > y)}_{y > k: = 0; y < k: \leq P(|X_k| > y)} dy \quad (5)$$

$$\leq \int_0^k 2yP(|X_1| > y)dy \quad (6)$$

Then we have

$$\sum_{k=1}^\infty \frac{\text{Var}(Y_k)}{k^2} \leq \sum_{k=1}^\infty \frac{1}{k^2} \int_0^\infty \mathbf{1}_{y < k} 2yP(|X_1| > y)dy \quad (7)$$

$$= \int_0^\infty \underbrace{\left(\sum_{k=1}^\infty \frac{1}{k^2} \mathbf{1}_{y < k} 2y\right)}_{A(y)} P(|X_1| > y)dy \quad (8)$$

$$\stackrel{\text{lemma c}}{\leq} 4 \int_0^\infty P(|X_1| > y)dy = 4E|X_1| \quad (9)$$

□

Lemma 2.27 (c.). $A(y) \leq 4 \quad \forall y \geq 0$

c.

$$\sum_{k \geq 0} \frac{1}{k^2} \leq \int_{m-1}^\infty \frac{1}{x^2} dx \quad (10)$$

$$= \frac{1}{m-1} \quad \text{for } m \geq 2, m \in \mathbb{Z} \quad (11)$$

For $y \geq 1$:

$$2y \sum_{k > y} \frac{1}{k^2} = 2y \sum_{k \geq [y] + 1} \frac{1}{k^2} \quad (12)$$

$$\leq \frac{2y}{[y]} \quad (13)$$

$$\leq 2 \frac{\lfloor y \rfloor + 1}{\lfloor y \rfloor} \quad (14)$$

$$= 2(1 + \frac{1}{\lfloor y \rfloor}) \quad (15)$$

$$\leq 4 \quad (16)$$

For $0 < y < 1$:

$$2y \sum_{k>y} \frac{1}{k^2} = 2y \sum_{k=1}^{\infty} \frac{1}{k^2} \quad (17)$$

$$\leq 2(1 + \sum_{k=2}^{\infty} \frac{1}{k^2}) \leq 4 \quad (18)$$

Remark. For $x \in \mathbb{R}$, $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}$

□

Let $\alpha > 1$ (real), $k(n) = \lfloor \alpha^n \rfloor$

$$\begin{aligned} \sum_{n=1}^{\infty} P(|T_{k(n)} - ET_{k(n)}| \geq \epsilon k(n)) &\leq \sum_{n=1}^{\infty} \frac{\text{Var}(T_{k(n)})}{\epsilon^2 k(n)^2} \\ &= \epsilon^{-2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{m=1}^{k(n)} \text{Var}(Y_m) \\ &= \epsilon^{-2} \sum_{m=1}^{\infty} \text{Var}(Y_m) \underbrace{\sum_{n:k(n) \geq m} \frac{1}{k(n)^2}}_{[*]} \\ &\leq \frac{4\epsilon^{-2}}{1 - \alpha^{-2}} \sum_{m=1}^{\infty} \frac{\text{Var}(Y_m)}{m^2} < \infty \quad \text{by lemma}(b) \end{aligned}$$

And we have

$$\begin{aligned} [*] &= \sum_{n:k(n) \geq m} \frac{1}{k(n)^2} \\ &= \sum_{n:\alpha^n \geq m} \frac{1}{[\alpha^n]^2} \\ &\leq 4 \sum_{n:\alpha^n \geq m} \alpha^{-2n} \\ &\leq 4 \frac{m^{-2}}{1 - \alpha^{-2}} \end{aligned}$$

Claim 2.28. $\lfloor \alpha^n \rfloor \geq \frac{\alpha^n}{2}$ because $\alpha > 1$

By B-C:

$$\frac{T_{k(n)}}{k(n)} - \frac{ET_{k(n)}}{k(n)} \xrightarrow{a.s.} 0$$

Check the $\frac{ET_{k(n)}}{k(n)}$:

$$EY_k = E[X_1 \cdot \mathbf{1}_{|X_1| \leq k}] \xrightarrow{k \rightarrow \infty} EX_1$$

by DCT and assumption $E|X_1| < \infty$

Thus

$$\frac{ET_{k(n)}}{k(n)} = \frac{1}{k(n)} \sum_{m=1}^{k(n)} E(Y_m) \xrightarrow{n \rightarrow \infty} EX_1$$

Remark. Cesaro convergence: $a_k \xrightarrow{k \rightarrow \infty} a \implies \frac{1}{n} \sum_{k=1}^n a_k \xrightarrow{n \rightarrow \infty} a$

Now we have:

$$\frac{T_{k(n)}}{k(n)} \xrightarrow{a.s.} EX_1$$

It's enough to prove the theorem for $X_k \geq 0$!

Because then

$$\begin{aligned} \frac{S_n}{n} &= \frac{1}{n} \sum_{k=1}^n X_k \\ &= \frac{1}{n} \sum_{k=1}^n X_k^+ - \frac{1}{n} \sum_{k=1}^n X_k^- \\ &\longrightarrow E(X_1^+) - E(X_1^-) = E(X_1) \end{aligned}$$

Now assume $X_k \geq 0$. Then also $Y_k \geq 0$.

Let $k(n) \leq m \leq k(n+1)$. $T_{k(n)} \leq T_m \leq T_{k(n+1)}$

$$\begin{aligned} \implies \frac{T_{k(n)}}{k(n+1)} &\leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)} \\ &\implies \frac{k(n)}{k(n+1)} \cdot \frac{T_{k(n)}}{k(n)} \\ &\leq \frac{T_m}{m} \\ &\leq \frac{k(n+1)}{k(n)} \cdot \frac{T_{k(n+1)}}{k(n+1)} \end{aligned}$$

Now we have $P\{\frac{EX_1}{\alpha} \leq \underline{\lim} \frac{T_m}{m} \leq \overline{\lim} \frac{T_m}{m}\} \leq \alpha EX_1 = 1$

Last step: $\alpha \searrow 1$ along some sequence. Then we get:

$$P\{EX_1 \leq \underline{\lim}_{m \rightarrow \infty} \leq \overline{\lim}_{m \rightarrow \infty} \frac{T_m}{m} \leq EX_1\} = 1$$

□

Application: Empirical distribution function:

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k \leq x} \xrightarrow[n \rightarrow \infty]{a.s.} E[\mathbf{1}_{X_1 \leq x}] \\ &= P(X_1 \leq x) = F(x) \end{aligned}$$

Theorem 2.29 (Glivenko-Cantelli theorem). For iid real-valued r.v.'s,

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \quad a.s.$$

Notice let $\epsilon > 0$, $x < y$: $F(y) - F(x) < \epsilon$, suppose we know: $|F(x) - G(x)| < \epsilon$, $|F(y) - G(y)| < \epsilon$.

For $z \in (x, y)$: $G(z) - F(z) \leq G(y) - F(x) = G(y) - F(y) + F(y) - F(x) < 2\epsilon$

3

[2.5] Let $\{X_k\}_{k=1}^\infty$ be r.v.'s on (ω, \mathcal{F}, P) .

$$\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}, \mathcal{F}'_n = \sigma\{X_n, \dots, X_{n+1}, \dots\}$$

We call $\mathcal{J} = \bigcap_{n=1}^\infty \mathcal{F}'_n$ be tail σ -algebra.

Example 3.1. 1. $\overline{\lim} X_n$ is \mathcal{T} -measurable.

$$\begin{aligned} \{\overline{\lim} X_n \geq a\} &= \bigcap_{k=1}^\infty \{X_n > a - \frac{1}{k} \text{ i.o.}\} \\ &= \bigcap_{k=1}^\infty \bigcap_{m=1}^\infty \bigcup_{n \geq m} \{X_n > a - \frac{1}{k}\} \\ &= \bigcap_{k=1}^\infty \bigcap_{m=l}^\infty \bigcup_{n \geq m} \{X_n > a - \frac{1}{k}\} \\ &\in \mathcal{F}'_l \quad \text{TRUE} \forall l \in \mathbb{Z}_{>0} \text{ (in) } \mathcal{T} \end{aligned}$$

Observation: B