Notes of Math 733: Probability Theory
Instructor: Timo Seppäläinen

YI WEI

Sep 2024

Contents

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 $\Omega, P(A) =$ the probability of the set (event) A

Axiom 1.1.

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

Example 1.1.

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

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

$$\begin{aligned}
\Omega &= \mathbb{Z}_{\geqslant 0} \\
\mathcal{F} &= \mathcal{P}(\Omega) \\
P(k) &= e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } k \in \Omega \\
P(A) &= \sum_{k \in A} e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } A \subseteq \Omega
\end{aligned}$$

3. Choose uniformly random real number from $[0, 1]$
   $P(x) = 0 \quad \forall\, x \in [0, 1]$
   if $0 \leqslant a < b \leqslant 1$:

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

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

$$\begin{aligned}
&\Omega = \{0, 1\}^{\mathbb{Z}_{\geqslant 0}} \\
&P\{w : x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n\} = 2^{-n} \qquad (*)
\end{aligned}$$

   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. $\varnothing \in \mathcal{A}$
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 <u>measurable space</u>.

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

1. $P(\varnothing) = 0$
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 <u>measure space</u>.

Definition 1.3. If X is a metric space, its <u>Borel σ-algebra</u> $\mathcal{B}_X$ 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

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

Definition 1.5. A probability space $(\Omega,\mathcal{F},P)$ 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 = samllest σ -algebra that contains all sets of the type

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

P obtained from Eq. *

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

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

A random variable X is a measurable function:

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

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:

$$\begin{aligned}X_1(w)&=x_1\quad\text{outcome of the first flip}\\X_2(w)&=x_2\quad\text{outcome of the second flip}\end{aligned}$$

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:

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

[**Date:** Sep 10,2024]

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

   Exercise 1.2. Check v 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 $\lambda(A)=0\Longrightarrow\mu(A)=0\quad\forall A\in\mathcal{A}$.

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

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

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

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

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

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

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

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

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

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

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

Then we know $\mu \ll$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 $(\Omega, \mathcal{F}, P)$

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.

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

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

Theorem 1.2.

Key result:

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

Here: X is a r.v. on $(\Omega, \mathcal{F}, P)$, $\mu = P \circ X^{-1}$ = distribution of X , $f : \mathbb{R} \to \mathbb{R}$ 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: $\int_{\Omega} \mathbf{1}_B(X(w))P(\mathrm{d}w)$ (same as $dP(w)$)

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

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

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

3. $f \geqslant 0, \exists$ simple function $0 \leqslant f_n$

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

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

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

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

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

□

Example 1.5. 1. $X \sim Possion(\lambda)$, $\mu =$ distribution of X. We know $\mu(B) = \sum_{k:k\in B} e^{-\lambda} \frac{\lambda^k}{k!} \Longrightarrow$ $\mu(\mathbb{R}\backslash\mathbb{Z}_{\geqslant 0}) = 0$. Then we have:

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

2. $X \sim Exp(\lambda)$

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

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 $\Omega, \mathcal{F}, P$ be a probability space.

Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ 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 $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if $\forall A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \ldots, A_n \in \mathcal{F}_n$, then

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

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

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

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

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

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

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 $P(A|Y = m) = \frac{P(A \bigcap \{Y=m\})}{P(Y=m)}$

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 and independent, we have:

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

Theorem 2.2. Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be subcollection of $\mathcal{F}$, Assume that each $\mathcal{A}_i$ is closed under intersection, which means ($A, B \in \mathcal{A}_i \implies A \bigcap B \in \mathcal{A}_i$) and $\Omega \in \mathcal{A}_i$. Assume that the probability $P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) \quad \forall A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n$. Then the σ-algebra $\sigma(\mathcal{A}_1), \ldots, \sigma(\mathcal{A}_n)$ 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: <u>Dynkin's $\pi - \lambda$ theorem</u>.

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 $A \subseteq B \Longrightarrow B \backslash A \in \mathcal{A}$
   (c) If $A_1 \subseteq A_2 \cdots \subseteq A_n \subseteq \cdots$ 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 $A_2 \in \mathcal{A}_2, \ldots, A_n \in \mathcal{A}_n$, set $\mathcal{F} = A_2 \bigcap \cdots \bigcap A_n$

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

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

Proof of Claim 2.4.

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

Take $A_1 = \Omega$

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

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 $B \backslash A \in \mathcal{L}$.

$$P\Big((B \backslash A) \bigcap F\Big) = P\Big((B \bigcap F) \backslash (A \bigcap F)\Big) = P(B \bigcap F) - P(A \bigcap F)$$

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

$$P(A \bigcap F) = \lim_{n \to \infty} P(A_n \bigcap F) \qquad \text{because } A_n \bigcap F \nearrow A \bigcap F$$

We've checked that $\mathcal{L}$ is a λ-system. So $\sigma(A_1) \subseteq \mathcal{L}$

$\square$

We continue the proof of theorem 2.2:

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

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 $\sigma(\mathcal{A}_1), \ldots, \sigma(\mathcal{A}_n)$ $\square$

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

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

[**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 $(X_1, \mathcal{A}_1, \mu_1), \ldots, (X_n, \mathcal{A}_n, \mu_n)$ are σ-finite measure spaces. The product measure space $(X, \mathcal{A}, \mu)$ is defined as follows:

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

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

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

Theorem 2.7 (Tonelli-Fubini Theorem)**.**

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

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 $d(f, g) = \sup_{0 \leqslant x < \infty} |f(x) - g(x)|$ makes not separable!
   But $d(f, g) = \sum_{n=1}^{\infty} 2^{-n} (\sup_{0 \leqslant x \leqslant n} |f(x) - g(x)| \wedge 1)$

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 $X_1, \ldots, X_n$ are r.v.'s on $(\Omega, \mathcal{F}, P)$. 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

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

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

Proof.

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

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

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

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

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 $\mu_i = P \circ X_i^{-1}$

$$\begin{aligned} E[f_1(X_i)f_2(X_2)] &= \int_{\mathbb{R}^2} f_1(x_1)f_2(x_2)(\mu_1 \otimes \mu_2)(dx_1 dx_2) \\ &= \int_R \mu_2(dx_2) \int_{\mathbb{R}} \mu_1(dx_1)f_1(x_1)f_2(x_2) \\ &= \int_{\mathbb{R}} u_2(dx_2)f_2(x_2) \int_{\mathbb{R}} u_1(x_1)f_1(x_1) \\ &= E\Big[f_1(X_1)\Big]E\Big[f_2(X_2)\Big] \end{aligned}$$

□

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

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

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

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

Definition 2.6 (convolutions)**.** Let μ, v 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 $\Longrightarrow$ the intepretation over y leaves a measurable function of the variable x.

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

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

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

Theorem 2.12. *$X \perp\!\!\!\perp Y$, $X \sim \mu$, $Y \sim \mu \Longrightarrow X + Y \sim \mu \times \nu$*

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

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

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

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

Finite case: Given $\mu_1, \dots, \mu_n$ Borel probability measure on $\mathbb{R}$.

Want: independent r.v,'s $X_1, \dots, X_n$ with $X_i \sim \mu_1$

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

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

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

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

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

Let $\Omega = \mathbb{R}^{\mathbb{Z}_{\geqslant 0}} = \{\omega = (x_i)_{i=1}^\infty : \text{ each } x_i \in \mathbb{R}\}$, $\mathcal{F} =$ product σ-algebra $= \sigma\{A_1 \times \dots \times A_n \times \mathbb{R} \times \mathbb{R} \dots : n \in \mathbb{Z}_{>0}, A_1, \dots, A_n \in \mathcal{B}_{\mathbb{R}}\} = \sigma$-algebra generated by the projection mapping $X_i(\omega) = x_i,\ i \in \mathbb{Z}_{>0},$ = smallest σ-algebra on Ω under which each $X_i : \Omega \to \mathbb{R}$ 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 $\{\mathbf{u}_n\}_{n \geqslant 1}$ 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 $\underbrace{\text{complete separable metric spaces.}}_{\text{"Polish spaces"}}$.

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 Durett)

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 $\sigma^2 = \mathrm{Var}(X_1) < \infty$, then

$$P\{\frac{S_n - n\mathbb{E}X_1}{\sigma\sqrt{n}} \leqslant s\} \xrightarrow{n\to\infty} \int_{-\infty}^{s} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}\,dx$$

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

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

Theorem 2.17 (1st Borel-Cantelli Lemmas)**.**

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

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

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

☐

(2.)

$$P(\bigcap_{m\geqslant 1}\bigcup_{n\geqslant m} A_n) = \lim_{m\to\infty} P(\bigcup_{n\geqslant m} A_n)$$
$$\leqslant \lim_{n\to\infty}\sum_{n\geqslant m} P(A_n) = 0 \quad \text{by convergent series tails}$$

□

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

Lemma 2.18. Suppose $\forall \epsilon > 0$, $\sum_{n=1}^{\infty} P(|x_n - x| \geqslant \epsilon) < \infty$.
Then $X_n \longrightarrow X \quad a.s.$.

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

$$\text{B-C} \Longrightarrow P(\bigcap_{m\geqslant 1}\bigcup_{n\geqslant m}\{|X_n - X| \geqslant \epsilon_j\}) = 0$$
$$\Longrightarrow 1 = P(\bigcup_{m\geqslant 1}\bigcap_{n\geqslant m}\{|X_n - X| < \epsilon_j\})$$
$$= P\{\exists m < \infty \ s.t. \ n \geqslant m \Longrightarrow |X_n - X| < \epsilon_j\}$$
$$1 = P(\bigcap_{j=1}^{\infty}\bigcup_{m\geqslant 1}\bigcap_{n\geqslant m}\{|X_n - X| < \epsilon_j\})$$
$$= P\{\forall j \ \exists m, \ m, n \geqslant m \Longrightarrow |X_n - X| < \epsilon_j\}$$
$$= P\{X_n \to X\}$$

□

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

Proof. $\sum_{n\geqslant 1} P(|Y_n| \geqslant \epsilon) \leqslant \sum_{n=1}^{\infty} \frac{E|Y_n|}{\epsilon} \leqslant \frac{1}{\epsilon}\sum_{n=1}^{\infty} 2^{-n} < \infty$ □

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

$$P(Z \geqslant a) = \mathbb{E}[\mathbf{1}_{Z\geqslant a}] \leqslant \mathbb{E}[\frac{Z}{a}\mathbf{1}_{Z\geqslant a}] \overset{Z\geqslant 0}{\leqslant} \mathbb{E}[\frac{Z}{a}]$$

Lemma 2.19 (Borel-Cantelli)**.** $\{A_n\}_{n\geqslant 1}$ a sequence of events on $(\Omega, \mathcal{F}, P)$ and $\sum_n P(A_n) < \infty \Longrightarrow P(\bigcap_{m=1}^{\infty}\bigcup_{n\geqslant m} A_n) = 0$

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

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

Proof. Let $M < N$.

$$P(\bigcap_{n=M}^{N} A_n^c) = \prod_{n=m}^{N}(1 - P(A_n)) \leqslant \prod_{n=M}^{N} e^{-P(A_n)}$$
$$= e^{-\sum_{n=M}^{N} P(A_n)} \underset{N\to\infty}{0}$$
$$\Longrightarrow P(\bigcap_{n=M}^{\infty} A_n^c) = 0$$
$$\Longrightarrow 1 = P(\bigcup_{n=M}^{\infty} A_n) = 1 \quad \text{true } \forall M$$

So

$$P(\bigcap_{M=1}^{\infty}\bigcup_{n=M}^{\infty}) = 1 \tag{2}$$

□

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

$$\sum_{n=1}^{\infty} P(X_n = 6) = \sum_{n=1}^{\infty} \frac{1}{6} = \infty$$
$$\overset{2\text{nd}B-C}{\Longrightarrow} P(X_n = 6 \text{ i.o.}) = 1$$

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

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

$$P(|X_n - X| \geqslant \epsilon) \underset{n\to\infty}{0}$$

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

$$1 = P(X_n \longrightarrow X) = P(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\{|X_n - X| \leqslant \frac{1}{k}\})$$
$$\leqslant P(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\{|X_n - X| \leqslant \frac{1}{k}\}) \quad \forall\, k \in \mathbb{Z}_{>0} \quad (\text{Notice that this increases with })m$$
$$= \lim_{m\to\infty} P(\bigcap_{n=m}^{\infty}\{|X_n - X| \leqslant \frac{1}{k}\})$$
$$\leqslant \underbrace{\lim_{m\to\infty} P(|X_m - X| \leqslant \frac{1}{k})}_{\text{this limit}=1\ \forall k}$$

Theorem 2.21.

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

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

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

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

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

- ($\Longleftarrow$) pf by contradiction: suppose $X_n \xrightarrow{P}$ fails. Then $\exists\epsilon > 0$ s.t. $P(|X_n - X| \geqslant \epsilon) \longrightarrow 0$ Fails. So $\exists$subseq $n(m)$ and $\delta > 0$ s.t. $P(|X_{n(m)} - X| \geqslant \epsilon) \geqslant \delta$. But by assumption $\exists$further subsequence converges a.s. $X_{n(m_k)} \overrightarrow{X}\, X \Longrightarrow X_{n(m_k)} \xrightarrow{P} X \Longrightarrow P(|X_{n(m_k)} - X| \geqslant \epsilon) \longrightarrow 0$

$\square$

Theorem 2.22 (SLLN and WLLN)**.** SLLN:

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

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

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

Lemma 2.23 (Chebyshev's Inequality)**.**

$$\begin{aligned} P(|X - \mathbb{E}X| \geqslant a) = P(|X - \mathbb{E}X|^2 \geqslant a^2) \\ \leqslant \frac{\mathbb{E}[|X - \mathbb{E}X|^2]}{a^2} \\ = \frac{\mathrm{Var}(X)}{a^2} \end{aligned}$$

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 <u>uncorrelated</u> is $0 = \mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = E[XY] - \mathbb{E}X\mathbb{E}Y$.

We know that $X \perp\!\!\!\perp Y \Longrightarrow X$ and Y uncorrelated.
Computation:

$$\mathrm{Var}(X + Y) = \mathbb{E}[(X + Y)^2] - (\mathbb{E}X + \mathbb{E}Y)^2$$
$$= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - (\mathbb{E}X)^2 - 2\mathbb{E}X\mathbb{E}Y - (\mathbb{E}Y)^2 = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$$

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 $Y_k = X_k \cdot \mathbf{1}_{\{|X_k| \leqslant k\}}$. 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.)

$$\begin{aligned}
\sum_{k=1}^{\infty} P(X_k \neq Y_k) &= \sum_{k=1}^{\infty} P(|X_k| > k) \\
&= \sum_{k=1}^{\infty} \int_{k-1}^{k} P(|X_k| > k)\mathrm{d}t \\
&\leqslant \sum_{k=1}^{\infty} \int_{k-1}^{k} P(|X_1|.t)\, dt \\
&= \int_{0}^{\infty} P(|X_1| > t)\, dt \\
&= E|X_1| \leqslant \infty
\end{aligned}$$

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 $\Longrightarrow$ $\exists$ random k_0 s.t.

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

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

$$\begin{aligned}
\frac{S_n}{n} &= \frac{T_n}{n} + \frac{S_n - T_n}{n} \\
\Longrightarrow \frac{|S_n - T_n|}{n} &\leqslant \frac{1}{n}\sum_{k=1}^{n} |X_k - Y_k| \\
\leqslant \frac{1}{n}\sum_{k=1}^{k_0(\omega)} |X_k(\omega) - Y_k(\omega)| &\underset{n\longrightarrow\infty}{\longrightarrow} 0
\end{aligned}$$

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

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

$$E[h(x)] = E[h(x) - h(0)] = E[\int_0^X h'(x)\mathrm{d}x] = E[\int_0^\infty h'(x)\mathbf{1}_{X>0}\mathrm{d}x] \tag{3}$$

$$= \int_0^\infty h'(x)P(X > 0)\mathrm{d}x \tag{4}$$

$\square$

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

Proof.

$$\mathrm{Var}(Y_k) \leqslant E(Y_k^2) = \int_0^\infty 2y \underbrace{P(|Y_k| > y)}_{y>k:=0;\ y<k:\leqslant P(|X_k|>y)} \mathrm{d}y \tag{5}$$

$$\leqslant \int_0^k 2yP(|X_1| > y)\mathrm{d}y \tag{6}$$

Then we have

$$\sum_{k=1}^\infty \frac{\mathrm{Var}(Y_k)}{k^2} \leqslant \sum_{k=1}^\infty \frac{1}{k^2}\int_0^\infty \mathbf{1}_{y<k}2yP(|X_1| > y)\mathrm{d}y \tag{7}$$

$$= \int_0^\infty \underbrace{(\sum_{k=1}^\infty \frac{1}{k^2}\mathbf{1}_{y<k}2y)}_{A(y)} P(|X_1| > y)\mathrm{d}y \tag{8}$$

$$\overset{\text{lemma c}}{\leqslant} 4\int_0^\infty P(|X_1| > y)\mathrm{d}y = 4E|X_1| \tag{9}$$

$\square$

Lemma 2.27 (c.). $A(y) \leqslant 4 \quad \forall\, y \geqslant 0$

c.

$$\sum_{k\geqslant 0} \frac{1}{k^2} \leqslant \int_{m-1}^\infty \frac{1}{x^2}\mathrm{d}x \tag{10}$$

$$= \frac{1}{m-1} \quad \text{for } m \geqslant 2,\ m \in \mathbb{Z} \tag{11}$$

For $y \geqslant 1$:

$$2y\sum_{k>y}\frac{1}{k^2} = 2y\sum_{k\geqslant[y]+1}\frac{1}{k^2} \tag{12}$$

$$\leqslant \frac{2y}{[y]} \tag{13}$$

$$\leqslant 2\frac{\lfloor y\rfloor + 1}{\lfloor y\rfloor} \tag{14}$$

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

$$\leqslant 4 \tag{16}$$

For $0 < y < 1$:

$$2y\sum_{k>y}\frac{1}{k^2} = 2y\sum_{k=1}^{\infty}\frac{1}{k^2} \tag{17}$$

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

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

□

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

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

And we have

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

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

By B-C:

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

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

$$EY_k = E[X_1 \cdot \mathbf{1}_{|X_1| \leqslant k}] \underset{k \longrightarrow \infty}{\longrightarrow} EX_1$$

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

Thus

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

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

Now we have:

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

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

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 $X_k \geqslant 0$. Then also $Y_k \geqslant 0$.

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

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

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

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

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

$\square$

Application: Empirical distribution function:

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \leqslant x} \xrightarrow[n\to\infty]{a.s.} E[\mathbf{1}_{X_1\leqslant x}]$$
$$= P(X_1 \leqslant x) = F(x)$$

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

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

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

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

2.3

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

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

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

Example 2.7. 1. $\overline{\lim}X_n$ is $\mathcal{T}$-measurable.

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

Observation: B

[**Date:** Oct 1, 2024]
Now how about $\bar{S} = \overline{\lim}_{n\to\infty}S_n$?

Suppose X_1 is something, $X_2 = X_3 = \cdots = X_n = \cdots = 0$

$$S_n = X_1 \implies \bar{S} = X_1 \text{ not } \mathcal{T}\text{-measurable, unless}$$

$X_k = 0,\ \forall k \implies \bar{S} = 0$ is $\mathcal{T}$-measurable.

Assume each X_k $\mathbb{R}$-valued. What about the event $\{\lim_{n\to\infty}$ exists in $\mathbb{R}\}$?

This is tail measurable:

Proof.

$$\bigcap_{k=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{m,n\geqslant N}\{|S_m - S_n| < \frac{1}{k}\}$$
$$\bigcap_{k=1}^{\infty}\bigcup_{N\geqslant l}^{\infty}\bigcap_{m,n\geqslant N}\{|S_m - S_n| < \frac{1}{k}\}$$
$$\iff \{|\sum_{i=N}^{m} X_i - \sum_{i=N}^{n} X_i| < \frac{1}{k}\} \in \mathcal{F}_l' \quad \forall l$$

□

Now we have $\eta_x : x \in \mathbb{Z}^2$. Define $\mathcal{F}_n = \sigma\{\eta_x : x \in [-n,n]^2 \bigcap \mathbb{Z}^2\}$. And $\mathcal{F}_n' = \sigma\{\eta_x : x \notin [-n,n]^2\}$, $\mathcal{T} = \bigcap_{n\geqslant 1}\mathcal{F}_n'$

Theorem 2.30 (Kolmogorov 0-1 law)**.** Let $\{X_k\}$ be independent. Then every $A \in \mathcal{T}$ satisfies $P(A) \in \{0,1\}$.

Proof. From the assumption and a $\pi - \lambda$ argument, F_n and $\mathcal{F}_{n+1}'$ are independent σ-algebras. (Thm.2.1.9), so if $A \in \mathcal{F}_n$ and $B \in \mathcal{T}$ are independent because the $\mathcal{T} \subseteq F_{n+1}'$.

Hence $\bigcup_{n\geqslant 1}\mathcal{F}_n$ and $\mathcal{T}$ are independent π-systems.

By a lemma we proved, $\sigma(\bigcup_{n\geqslant 1}\mathcal{F}_n)$ is independent of $\sigma(\mathcal{T}) = \mathcal{T}$. Let $A \in \mathcal{T}$. $T \subseteq \mathcal{F}_n' = \sigma\{X_n, \ldots, \}$ We have

$$\sigma(\bigcup_{n\geqslant 1}\mathcal{F}_n) \implies A \in \sigma(\bigcup n \geqslant 1).$$

So A is independent of itself! Lence $P(A) = P(A \bigcap A) = (P(A))^2 \implies P(A) \in \{0,1\}$ □

Example 2.8. An X_n independent. Then we know that $\{\bar{X} \geqslant 0\} \in \{0,1\} \forall a \in \mathbb{R}$.

Cliam: $\exists c \in [-\infty, \infty]$ s.t. $F(X = c) = 1$

Proof. Case 1: $P(\bar{X} > a) = 0 \quad a \in \mathbb{R}$.

$$P(\bar{X} < \infty = P(\bigcap_{k=1}^{\infty})) = 1$$

Case 2: $\exists a \in \mathbb{R}$ s.t. $P(\bar{X} > a) = 1$. Lt $C = \sup\{a \in \mathbb{R} : P(\bar{X}) > 0\}$ [**YW:** To update it] □

3 Weak convergence

Suppose we want to define what is it means for propabilityu on $\mathbb{R}$ to converge.

Possible def: $\mu_n \to$ if $\mu_n(A) \underset{n\to\infty}{\longrightarrow} \quad \forall A \in B_{\mathbb{R}}$

Example 3.1. 1. $\mu_n = S_{\frac{1}{n}}, \frac{1}{n} \underset{n\to\infty}{\longrightarrow}$ so we might want $S_{\frac{1}{n}} \underset{n\to\infty}{\longrightarrow} S_0$. $[S_x(A)] = \mathbf{1}_A(x)$, but $S_{\frac{1}{n}}(\{0\}) = 0$ and $S_0(\{0\}) = 1$.

We abandon the above definition and we have:

Definition 3.1. Let (S, ϱ) be a metric space with its Borel σ-algebra B_S. Let $C_b(S) = \{f : S \to \mathbb{R}, f$ is continuous and $\exists$ constant $C s.t.\} |f(x)| \leqslant C \quad \forall x \in S$.

Let $\{\mu_n\}_{n\geqslant 1}$ and μ be Borel probability measure on S. Then $\mu_n \longrightarrow \mu$ weakly if $\forall f \in C_b(S)$:

$$\int_S f\, d\mu_n \underset{n\to\infty}{\longrightarrow} \int_S f\, du$$

Example 3.2. $\int f\, d\delta_{\frac{1}{n}} = f(\frac{1}{n}) \longrightarrow f(0) = \int f d\delta_0$, so

$$\delta_{\frac{1}{n}} \longrightarrow \delta_0 \text{ weakly.}$$

Alternative notation:

$$\mu_n \overset{w}{\mu},\ \mu_n \overset{d}{\longrightarrow},\ \mu_n \Longrightarrow \mu$$

If $X_n \sim \mu_n, X \sim \mu$ and $\mu_n \overset{w}{\longrightarrow} \mu$ then we write $X_n \overset{w}{argument} X$, $X_n \overset{d}{X}$, $X_n \Longrightarrow X$.

Theorem 3.1 (Portmanteam theorem)**.** Let μ_n, μ be Borel. Then the following are equivalent:

1. $\int f d\mu_n \longrightarrow \int f d\mu \quad \forall f \in C_b(S)$.
2. $\int f d\mu_n \longrightarrow \int f d\mu$ for all bounded Lipschitz functions $f : S \to \mathbb{R}$. [Def: f is Lipschitz if there exists a constant L s.t. $|f(x) - f(y)| \leqslant L\varrho(x, y)$] Lipschitz functions are almost everywhere differentiable.
3. For closed sets $F \subseteq S$, $\overline{\lim}_{n\to\infty} \mu_n(F) \leqslant \mu(F)$.
4. For open sets $G \subseteq S$, $\underline{\lim}_{n\to\infty} \mu_n G \geqslant \mu(G)$.
5. If $A \in \mathcal{B}_S$ and $\mu(\partial A) = 0$, then $\mu_n(A) \longrightarrow \mu(A)$.
6. If $f : S \to \mathbb{R}$ is a bounded Borel function, and $\mu(D_f) = 0$ for $D_f = \{x \in S : f$ is discontinuous at $x\}$, then

$$\int d d\mu_n \longrightarrow \int f d\mu$$

Remark. $\partial A = BdA = \bar{A} - A^0$

Example 3.3. continue the example above: $\int f\, d\delta_{\frac{1}{n}} = f(\frac{1}{n}) \longrightarrow f(0) = \int f d\delta_0$, so

$$\delta_{\frac{1}{n}} \longrightarrow \delta_0 \text{ weakly.}$$

Proof.

(1) $\implies$ (2) needs no proof

(2) $\implies$ (3) Distance between a point and a set is $dist(x, F) = inf\{\varrho(x, z) : z \in F\}$.

$$f_k(x) = (1 - k \cdot dist(x, F))^+$$

Here f_k is Lipschitz.

Remark. Key Fact: F closed $\implies [x \in F \iff dist(x, F) = 0]$

$\mathbf{1}_F \leqslant f \leqslant 1$ and $f_k(x) \searrow \mathbf{1}_F(x)$ as $k \nearrow \infty$.

$$\overline{\lim_{n\to\infty}} \mu_n(F) \leqslant \overline{\lim_{n\to\infty}} \inf f_k d\mu_n = \int f_k d\mu \; k \underset{k\to\infty}{\longrightarrow} \int \mathbf{1}_F d\mu$$

(3) $\implies$ (4) by taking complements.

$$\overline{\lim} \mu_n(F) \leqslant \mu(F) \iff \underline{\lim} \mu_n(F^c) \geqslant \mu(F^c)$$

(3 & 4) $\implies$ (5)

$$\begin{aligned} \overline{\lim} \mu_n(A) &\leqslant \overline{\lim} \mu_n \bar{A} \leqslant \mu(\bar{A}) \\ &\leqslant \mu(A) + \mu(\partial A) = \mu(A) \end{aligned}$$

Now we check the other direction:

$$\underline{\lim} \mu_n(A) \geqslant \underline{\lim} \mu_n(A^0) \geqslant \mu(A^0) \geqslant \mu(\bar{A}) - \mu(\partial A) \geqslant \mu(A)$$

$\square$