Igor Kortchemski
Probality theory,
Autumn 2023
ETH

Chapter 2: random variables

Outline

1) Measurable functions
2) Product σ-fields and families of functions
3) Independent random variables
4) Real-valued random variables
5) Integration
6) Classical laws
7) Integration and independence

1) Measurable functions

Definition Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces. A function $f:(E,\mathcal{E})\to(F,\mathcal{F})$ is measurable if $\forall B\in\mathcal{F}$, $f^{-1}(B)\in\mathcal{E}$.

Remarks:
- Since $(f\circ g)^{-1}(B)=g^{-1}(f^{-1}(B))$, a composition of measurable functions is measurable
- If $\mathcal{E}'$ is a σ-field with $\mathcal{E}\subset\mathcal{E}'$, if $f:(E,\mathcal{E})\to(F,\mathcal{F})$ is measurable, then $f:(E,\mathcal{E}')\to(F,\mathcal{F})$ is measurable

Criterion To check that $f:(E,\mathcal{E})\to(F,\mathcal{F})$ is measurable, one often finds a class $\mathcal{C}\subset\mathcal{F}$ such that $\mathcal{F}=\sigma(\mathcal{C})$ and $\forall B\in\mathcal{C}$, $f^{-1}(B)\in\mathcal{E}$.
Indeed, $\{B\in\mathcal{F} ; f^{-1}(B)\in\mathcal{E}\}$ is then a σ-field (exercise) containing $\mathcal{C}$ and thus containing $\sigma(\mathcal{C})=\mathcal{F}$.

Interpretation in probability. A measurable function $X:(\Omega,\mathcal{A})\to(F,\mathcal{F})$ is called a random variable (r.v. in short). Intuitively, this means that $X(\omega)$ is "observable" in the sense one can "observe" whether $X(\omega)\in B$ with $B\in\mathcal{F}$ or not.

Definition Let $f:(E,\mathcal{E})\to(F,\mathcal{F})$ be a measurable function and let μ be a measure on $(E,\mathcal{E})$. Then, $\forall B\in\mathcal{F}$, $\mu_f(B)=\mu(f^{-1}(B))$ defines a measure on $(F,\mathcal{F})$, called the image measure of f by μ.

In probability, if $X:(\Omega,\mathcal{E})\to(F,\mathcal{F})$ is a random variable and $\mathbb{P}$ a probability measure on $(\Omega,\mathcal{E})$, then $\mathbb{P}_X$, the image measure of $\mathbb{P}$ by X is called the <u>law</u> of X.
By definition, $\forall B\in\mathcal{F}$,

$$\mathbb{P}_X(B)=\mathbb{P}(X^{-1}(B))=\mathbb{P}(\{\omega\in\Omega;\ X(\omega)\in B\})$$
$$=\mathbb{P}(X\in B)$$

probabilistic notation (avoiding writing "ω")

Side remark If $(E,\mathcal{E},\mu)$ is a probability space, there exists a random variable with law μ: just take $(\Omega,\mathcal{F},\mathbb{P})=(E,\mathcal{E},\mu)$ and $X:\Omega\to E$, $x\mapsto x$ the identity.

END OF LECTURE 4

How can one characterize a probability measure? In other words, is their a "simple" way to check if $\mathbb{P}_X=\mathbb{P}_Y$? (i.e $\mathbb{P}_X(A)=\mathbb{P}_Y(A)\ \forall A\in\mathcal{A}$)

If X takes it values in a countable space $(E,\mathcal{P}(E))$, then its law is characterized by the values $\mathbb{P}_X(\{x\})=\mathbb{P}_X(x)=\mathbb{P}(X=x)$. for $x\in E$, since for $A\subset E$, $\mathbb{P}_X(A)=\sum_{x\in A}\mathbb{P}_X(x)$

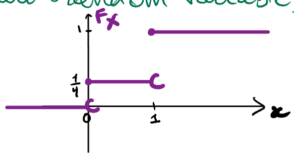
In particular, $\mathbb{P}(X=x)=\mathbb{P}(Y=x)\ \forall x\in E\Rightarrow\mathbb{P}_X=\mathbb{P}_Y$.

When E is uncountable this is not true anymore and things can be very complicated!

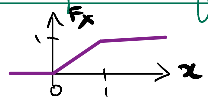
In the case of $\mathbb{R}$, cumulative distribution functions are useful:

Definition The <u>cumulative distribution function (cdf)</u> of a real-valued rv X is the function $F_X:\mathbb{R}\to[0,1]$ defined by $F_X(x)=\mathbb{P}(X\le x)=\mathbb{P}_X(]-\infty,x])$ for $x\in\mathbb{R}$

Example (Bernoulli random variable) If $\mathbb{P}(X=0)=\frac{1}{4}$ and $\mathbb{P}(X=1)=\frac{3}{4}$,



Example (Uniform distribution) If the law of X is the Lebesgue measure on $[0,1]$:



(X is said to follow the uniform distribution on $[0,1]$)

Proposition

① Let X be a $\mathbb{R}$-valued r.v. F_X is non-decreasing, $\lim_{-\infty} F_X = 0$, $\lim_{+\infty} F_X = 1$, F_X is right-continuous

② If X, Y are two $\mathbb{R}$-valued r.v, $F_X = F_Y \Leftrightarrow \mathbb{P}_X = \mathbb{P}_Y$ (cdf's characterize laws)

③ [Lebesgue-Stieltjes] If a function $F:\mathbb{R}\to[0,1]$ satisfies the condition in ①, then there exists real-valued r.v. with $F = F_X$.

Proof: ① • Since $]-\infty,x] \subset]-\infty,y]$ for $x\le y$, we get $F_X(x) = \mathbb{P}_X(]-\infty,x]) \le \mathbb{P}_X(]-\infty,y]) = F_X(y)$, so F_X is non-decreasing

• Since $\bigcap_{k=1}^{n}]-\infty,-k] =]-\infty,-n]$ is decreasing in n, we get $F_X(-n) \xrightarrow[n\to\infty]{} \mathbb{P}_X\left(\bigcap_{k=1}^{\infty}]-\infty,-k[\right) = \mathbb{P}_X(\emptyset) = 0$

Similarly, $\bigcup_{k=1}^{n}]-\infty,k] =]-\infty,n]$ is increasing in n, so $F_X(n) \xrightarrow[n\to\infty]{} \mathbb{P}_X\left(\bigcup_{k=1}^{\infty}]-\infty,k]\right) = \mathbb{P}_X(\mathbb{R}) = 1$.

• Take $x\in\mathbb{R}$. Since $\bigcap_{k=1}^{n}]-\infty, x+\frac{1}{k}[$ is decreasing in n, we get
$F_X(x+\frac{1}{n}) \xrightarrow[n\to\infty]{} \mathbb{P}_X\left(\bigcap_{k=1}^{\infty}]-\infty, x+\frac{1}{k}[\right) = \mathbb{P}_X(]-\infty,x]) = F_X(x)$

② • If $\mathbb{P}_X = \mathbb{P}_Y$, the $\forall x\in\mathbb{R}$, $\mathbb{P}_X(]-\infty,x]) = \mathbb{P}_Y(]-\infty,x])$

• If $F_X = F_Y$, then the probability measures $\mathbb{P}_X$ and $\mathbb{P}_Y$ coincide on $\{]-\infty,x], x\in\mathbb{R}\}$, which is a generating π-system of $\mathcal{B}(\mathbb{R})$

By the result of lecture 3, we conclude that $\mathbb{P}_X = \mathbb{P}_Y$.

③ Take $\Omega =]0,1[$, $\mathcal{A} = \mathcal{B}(]0,1[)$, λ Lebesgue measure on $]0,1[$. For $\omega\in]0,1[$ set $X(\omega) = \inf\{t\in\mathbb{R} : F(t)\ge\omega\}$ (called the right-continuous inverse of F). One checks that X is measurable and that $X(\omega)\le x \Leftrightarrow \omega\le F(x)$ for $\omega\in\Omega$ and $x\in\mathbb{R}$

Thus $\mathbb{P}(X\le x) = \mathbb{P}(\{\omega\in\Omega : \omega\le F(x)\}) = F(x)$.

Notation If $(F,\mathcal{F})$ is a measurable space, E a set and $f: E \to F$ a function, we define
$\sigma(f) = \{ f^{-1}(B) : B \in \mathcal{F} \}$ is a σ-field on E called the σ-field generated by f.
More generally, if $(f_i)_{i\in I}$ is a collection of functions with $f_i : E \to F_i$ and $(F_i, \mathcal{F}_i)$ measurable space we define $\sigma(f_i, i\in I) = \sigma\left(\{ f_i^{-1}(B_i) : i \in I, B_i \in \mathcal{F}_i \} \right)$, called the σ-field generated by $(f_i)_{i\in I}$

Remark In general, $\sigma(f_i : i \in I) \neq \bigcap_{i\in I} \sigma(f_i)$: we have $\sigma(f_i : i\in I) = \sigma\left(\bigcup_{i\in I} \sigma(f_i) \right)$ (see exercise sheet)

Example: If $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$, then $\sigma(f) = \{ A \in \mathcal{B}(\mathbb{R}) : A = -A \}$

Interpretation in probability

- $\sigma(X)$ represents the "information" / "observable sets" one has access to by looking at the values of X.
- Similarly, if $(X_i)_{i\in I}$ are r.v., $\sigma(X_i, i\in I)$ represents the "information" given by $(X_i)_{i\in I}$

Proposition ① Let $f : E \to (F, \mathcal{F})$ be a function. Then $\sigma(f)$ is the smallest σ-field on E for which f is measurable
② Let $f_i : E \to (F_i, \mathcal{F}_i)$ for $i \in I$ be functions. Then $\sigma(f_i : i \in I)$ is the smallest σ-field on E for which $\forall i \in I$ f_i is measurable

Proof ① • First $f : (E, \sigma(f)) \to (F, \mathcal{F})$ is measurable by definition of $\sigma(f)$
• If $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ we show that $\sigma(f) \subset \mathcal{E}$. To this end, by measurability of f, $\forall B \in \mathcal{F}$, $f^{-1}(B) \in \mathcal{E}$, so $\sigma(f) \subset \mathcal{E}$ by definition of $\sigma(f)$
② The proof is similar. First, for $i \in I$ $f_i : (E, \sigma(f_i : i\in I)) \to (F_i, \mathcal{F}_i)$ is measurable because for every $B_i \in \mathcal{F}_i$, $f_i^{-1}(B_i) \in \sigma(f_i : i \in I)$ by definition of $\sigma(f_i : i\in I)$
• If $f_i : (E, \mathcal{E}) \to (F_i, \mathcal{F}_i)$ is measurable $\forall i \in I$ we show that $\sigma(f_i : i \in I) \subset \mathcal{E}$. To this end, by measurability of f_i $\forall B_i \in \mathcal{F}_i$ $f_i^{-1}(B_i) \in \mathcal{E}$. Thus $\mathcal{E}$ contains all sets of the form $f_i^{-1}(B_i)$ with $i \in I$, $B_i \in \mathcal{F}_i$, so it also contains the σ-field they generate, which is precisely $\sigma(f_i : i \in I)$.

~

If E is a metric space, recall that $\mathcal{B}(E)$ is the Borel σ-algebra, generated by all open sets of E.

Proposition Let E, F be metric spaces and $f : (E, \mathcal{B}(E)) \to (F, \mathcal{B}(F))$ continuous. Then f is measurable.

Proof: $\forall O \in \mathcal{B}(F)$, if O is open, $f^{-1}(O)$ is open, hence is in $\mathcal{B}(E)$. Since $\mathcal{B}(F) = \sigma(\text{open sets of } F)$, this shows that $f^{-1}(B) \in \mathcal{B}(E)$ for every $B \in \mathcal{B}(F)$.

2) Product σ-fields and families of functions

Product σ-fields are needed when considering pairs, or more generally families, of random variables

Definition (product σ-field) Let $(E_i, \mathcal{E}_i)_{i \in I}$ be measurable spaces. Set $E = \prod_{i \in I} E_i$ and for $i \in I$, let $\pi_i : E \to E_i$ be the canonical projections.

We let $\bigotimes_{i \in I} \mathcal{E}_i := \sigma(\pi_i, i \in I) = \sigma\left(\pi_i^{-1}(B_i) : B_i \in \mathcal{E}_i \text{ for } i \in I\right)$ be the smallest σ-field on E such that all the projections are measurable. It is called the product σ-field or cylinder σ-field.

Definition (cylinder sets) Sets of the form

$$\pi_{i_1}^{-1}(A_1) \cap \cdots \cap \pi_{i_n}^{-1}(A_n)$$

with $i_1, \dots, i_n \in I$, $A_1 \in \mathcal{E}_{i_1}, \dots, A_n \in \mathcal{E}_{i_n}$ are called cylinder sets (they form a generating π-system of the product σ-field)

Examples. In $\mathbb{R}^{[0,1]} = \{f : [0,1] \to \mathbb{R}\}$, the set

$$\{f : [0,1] \to \mathbb{R},\ -1 < f(\tfrac{1}{2}) < 1 \text{ and } f(1) > \sqrt{2}\}$$
$$= \pi_{\frac{1}{2}}^{-1}(]-1,1[) \cap \pi_1^{-1}(]\sqrt{2}, \infty[)$$ is a cylinder set

- If $I = \{1, 2, \dots\}$, $E_i = \{0,1\}$, the product σ-field on $\prod_{i \in I} E_i = \{0,1\}^{\{1,2,\dots\}}$ is the cylinder σ-field seen in lecture 2

Proposition If $\#I=n$, $\bigotimes_{i=1}^{n} \mathcal{E}_i = \sigma(A_1\times\cdots\times A_n : A_i\in\mathcal{E}_i \text{ for } 1\le i\le n)$

Proof: Set $\mathcal{E}=\sigma(A_1\times\cdots\times A_n : A_i\in\mathcal{E}_i \text{ for } 1\le i\le n)$

- $\Pi_i : (E,\mathcal{E}) \to \mathcal{E}_i$ is measurable because $\Pi_i^{-1}(B_i) = E_1\times\cdots\times E_{i-1}\times B_i\times E_{i+1}\times\cdots\times E_n\in\mathcal{E}$
- if Π_i is measurable $\forall i$, then $A_1\times\cdots\times A_n = \Pi_1^{-1}(A_1)\cap\cdots\cap\Pi_n^{-1}(A_n)$

Hence $\mathcal{E}$ is the smallest σ-algebra s.t. $\forall i$, Π_i is measurable

END OF LECTURE 5

WARNING: It is not true in general that $\bigotimes_{i\in I}\mathcal{E}_i = \sigma(\prod_{i\in I} A_i : A_i\in\mathcal{E}_i)$ (it is true when I is countable, but not in general)

Remark If $\mathcal{C}_i$ is a generating π-system of $\mathcal{E}_i$, then $\{A_1\times\cdots\times A_n : A_i\in\mathcal{C}_i\}$ is a generating π-system of $\mathcal{E}_1\otimes\cdots\otimes\mathcal{E}_n$

Definition • If $\mathbb{P}_i$ is a probability measure on $(E_i,\mathcal{E}_i)$, the product probability measure $\bigotimes_{i\in I}\mathbb{P}_i$ is the unique probability measure on $(\prod_{i\in I}E_i, \bigotimes_{i\in I}\mathcal{E}_i)$ such that $\bigotimes_{i\in I}\mathbb{P}_i(\Pi_{i_1}^{-1}(A_1)\cap\cdots\cap\Pi_{i_k}^{-1}(A_k)) = \mathbb{P}_{i_1}(A_1)\times\cdots\times\mathbb{P}_{i_k}(A_k)$ for all cylinder sets.

• When $\#I=n$, $\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_n$ is the unique probability measure on $(E_1\times\cdots\times E_n, \mathcal{E}_1\otimes\cdots\otimes\mathcal{E}_n)$ such that $\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_n(A_1\times\cdots\times A_n) = \mathbb{P}_1(A_1)\,\mathbb{P}_2(A_2)\cdots\mathbb{P}_n(A_n)$ for $A_i\in\mathcal{E}_i$.

(uniqueness comes from the fact that 2 probability measures that coincide on a generating π-system are equal; existence is admitted: it involves additional tools from measure theory)

The definition extends to σ-finite measures such as the Lebesgue measure (μ is σ-finite on E if one can write $E=\bigcup_{n\ge1}E_n$ with $\mu(E_n)<\infty$).

Example The Lebesgue measure on $\mathbb{R}^n$ is the product $\lambda\otimes\cdots\otimes\lambda$ with λ the Lebesgue measure on $\mathbb{R}$.

Remark When one considers a family of random variables, their law is a probability measure on a product space. For example,
$\mathbb{P}(XY\le 1) = \mathbb{P}_{(X,Y)}(\{(x,y)\in\mathbb{R}^2 : xy\le 1\}) = \mathbb{P}_{XY}(]-\infty,1])$.

More generally, if $(X_1,\dots,X_n)$ is a random variable with values in $(E_1,\dots,E_n)$ its law $\mathbb{P}_{(X_1,\dots,X_n)}$ is characterized by $\mathbb{P}_{(X_1,\dots,X_n)}(A_1\times\cdots\times A_n) = \mathbb{P}((X_1,\dots,X_n)\in A_1\times\cdots\times A_n)$

$= \mathbb{P}(X_1\in A_1,\dots,X_n\in A_n)$

Proposition ① Let $(E_i, \mathcal{E}_i)_{i\in I}$ be a family of measurable spaces.
A function $f : (E, \mathcal{E}) \longrightarrow \left(\prod_{i\in I} E_i, \bigotimes_{i\in I} \mathcal{E}_i\right)$ is measurable iff $\forall i \in I$ $\pi_i \circ f$ is measurable (i.e writing $f = (f_i)_{i\in I}$, f is measurable iff $\forall i\in I$, f_i is measurable)

② If $f, g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are measurable, $f+g$, $f\times g$, $\min(f,g)$, $\max(f,g)$ are measurable

Thus, if $(X_i)_{i\in I}$ is a collection of random variables, one may view $(X_i)_{i\in I}$ as ONE random variable

Proof: ① $\boxed{\Rightarrow}$ If f is measurable, $\forall i\in I$ $\pi_i\circ f$ is a composition of measurable function, so it is measurable.

$\boxed{\Leftarrow}$ Since $\bigotimes_{i\in I} \mathcal{E}_i = \sigma(\pi_i^{-1}(B_i) : B_i \in \mathcal{E}_i, i\in I)$, it suffices to check that $f^{-1}(\pi_i^{-1}(B_i)) \in \mathcal{E}$ for every $i \in I$ and $B_i \in \mathcal{E}_i$.

But $f^{-1}(\pi_i^{-1}(B_i)) = (\pi_i \circ f)^{-1}(B_i) \in \mathcal{E}_i$ because $\pi_i \circ f$ is measurable

② By ①,
$$\begin{aligned} (\mathbb{R}, \mathcal{B}(\mathbb{R})) &\longrightarrow (\mathbb{R}^2, \mathcal{B}(\mathbb{R})\otimes \mathcal{B}(\mathbb{R})) \\ x &\longmapsto (f(x), g(x)) \end{aligned}$$
is measurable.

But
$$\begin{aligned} (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) &\longrightarrow \mathbb{R} \\ (u,v) &\longmapsto u+v \end{aligned}$$
is continuous, hence measurable

Since $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ (see exercise sheet), we conclude by composition that $f+g$ is measurable. the reasoning is the same for the other functions

3) Independent random variables

For a function $X : \Omega \to (E, \mathcal{E})$ recall that $\sigma(X) = \{X^{-1}(A) : A \in \mathcal{E}\}$

Definition ($\perp\!\!\!\perp$ of a finite number of random variables)
Random variables $X_1, \dots, X_n$ are <u>independent</u> (notation: $\perp\!\!\!\perp$) if the σ-fields $\sigma(X_1), \dots, \sigma(X_n)$ are independent.

Remark By the definition of independence, if $X_i : (\Omega, \mathcal{F}) \to (E_i, \mathcal{E}_i)$ are random variables,

$X_1, \dots, X_n$ are $\perp\!\!\!\perp$ $\iff$ $\forall A_1 \in \mathcal{E}_1, \dots, \forall A_n \in \mathcal{E}_n$, $\mathbb{P}(X_1^{-1}(A_1) \cap \dots \cap X_n^{-1}(A_n)) = \mathbb{P}(X_1^{-1}(A_1)) \dots \mathbb{P}(X_n^{-1}(A_n))$

$\iff \forall A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n \quad \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_n \in A_n)$

$\iff \forall A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n \quad \mathbb{P}_{(X_1, \dots, X_n)}(A_1 \times \dots \times A_n) = \mathbb{P}_{X_1}(A_1) \dots \mathbb{P}_{X_n}(A_n)$

$\underset{(*)}{\iff} \mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$

The last $\iff$ comes from the fact that two probability measures are equal iff they are equal on a generating π-system.

Remark To show that $X_1, \dots, X_n$ are $\perp\!\!\!\perp$ one very often shows that

$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_n \in A_n)$ $(*)$

for every $A_1 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$ with $\mathcal{C}_i$ a generating π-system of $\mathcal{E}_i$ with $\Omega \in \mathcal{C}_i$

(thanks to the property seen at the end of Chapter 1)

Proposition

(1) If $X_1, \dots, X_n$ are $\mathbb{Z}$-valued r.v, they are $\perp\!\!\!\perp$ iff $\forall i_1, \dots, i_n \in \mathbb{Z}$,

$$\mathbb{P}(X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_1 = i_1) \dots \mathbb{P}(X_n = i_n)$$

(2) If $X_1, \dots, X_n$ are $\mathbb{R}$-valued r.v, they are $\perp\!\!\!\perp$ iff $\forall x_1, \dots, x_n \in \mathbb{R}$,

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \dots \mathbb{P}(X_n \le x_n)$$

Proof: (1) Take $\{\mathbb{Z}\} \cup \{\{i\} : i \in \mathbb{Z}\}$, generating π-system of $\mathbb{Z}$. To see that $(*)$ holds with $A_k = \mathbb{Z}$, just sum over all $i_k \in \mathbb{Z}$ (for example, for $n=3$, $\mathbb{P}(X_1 = i_1, X_2 = i_2) = \sum_{i_3 \in \mathbb{Z}} \mathbb{P}(X_1 = i_1, X_2 = i_2, X_3 = i_3)$)

(2) Take $\{\mathbb{R}\} \cup \{]-\infty, x] : x \in \mathbb{R}\}$, generating π-system of $\mathbb{R}$. To see that $(*)$ holds with $A_k = \mathbb{R}$, just make $x_k \to \infty$.

Definition Let $X = (X_1, \dots, X_n)$ be a random variable in $E_1 \times \dots \times E_n$. The law $\mathbb{P}_{X_i}$ of X_i is called a marginal law. The law of $(X_1, \dots, X_n)$ is called the joint law.

Remark Since $\mathbb{P}_{(X_1, \dots, X_n)}(E_1 \times \dots \times E_{i-1} \times A_i \times E_{i+1} \times \dots \times E_n) = \mathbb{P}_{X_i}(A_i)$, the marginal laws are determined by the joint law $\mathbb{P}_{(X_1, \dots, X_n)}$. The converse is false in general, but it is true for independent random variables by $(*)$

Lemma (Composition principle) Let $(X_i)_{1\le i\le n}$ be random variables with $X_i : \Omega \to E_i$. Let $f_i : E_i \to F_i$ be measurable. Assume that $(X_i)_{1\le i\le n}$ are $\perp\!\!\!\perp$.
Then $(f_i(X_i))_{1\le i\le n}$ are $\perp\!\!\!\perp$

Proof: This follows from the fact that $\sigma(f_i(X_i)) \subset \sigma(X_i)$.
Indeed, take $A \in \sigma(f_i(X_i))$. Then $\exists B \in \mathcal{F}_i$ with $A = (f_i \circ X_i)^{-1}(B) = X_i^{-1}(f_i^{-1}(B))$
Since f_i is measurable, $f_i^{-1}(B) \in \mathcal{E}_i$, so $A \in \sigma(X_i)$.

Definition (independence for any family of random variables)
If $(X_i)_{i\in I}$ are random variables, they are independent (notation: $\perp\!\!\!\perp$) if $\forall J \subset I$, $\mathrm{Card}(J) < +\infty$, $(X_j)_{j\in J}$ are independent.

Let us give two useful results involving infinite families (countable for the first one, uncountable for the second one).

Lemma Let $(X_i)_{i\ge 1}$ be $\perp\!\!\!\perp$. Fix $p \ge 1$. Set
$\mathcal{B}_1 = \sigma(X_1, \dots, X_p)$ and $\mathcal{B}_2 = \sigma(X_k : k \ge p+1)$
Then $\mathcal{B}_1 \perp\!\!\!\perp \mathcal{B}_2$

Proof: We apply the previous proposition of lecture 4 with
$$\mathcal{C}_1 = \sigma(X_1, \dots, X_p) \quad \text{and} \quad \mathcal{C}_2 = \bigcup_{k=p+1}^{\infty} \sigma(X_{p+1}, \dots, X_k)$$

Indeed, $\sigma(\mathcal{C}_1) = \mathcal{B}_1$, $\sigma(\mathcal{C}_2) = \mathcal{B}_2$ (cf exercise sheet) $\mathcal{C}_1$ and $\mathcal{C}_2$ are π-systems such that $\forall A \in \mathcal{C}_1$, $\forall B \in \mathcal{C}_2$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
Indeed, we have $\sigma(X_1, \dots, X_p) \perp\!\!\!\perp \sigma(X_{p+1}, \dots, X_k)$ by the coalition principle for independent σ-fields.

END OF LECTURE 6

Remark The coalition principle for $\perp\!\!\!\perp$ σ-fields shows that if $(X_i)_{1\leq i\leq n}$ are $\perp\!\!\!\perp$ r.v
$(X_1,\dots,X_{n_1})$, $(X_{n_1+1},\dots,X_{n_2})$, $\dots$, $(X_{n_k+1},\dots,X_n)$ are $\perp\!\!\!\perp$
The previous proposition covers a case of an infinite family

Lemma The two random variables $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ with values in $\prod_{i\in I} E_i$ and $\prod_{j\in J} F_j$ are independent if and only if $\forall i_1,\dots,i_k\in I$ $\forall j_1,\dots,j_\ell\in J$,

$$(X_{i_1},\dots,X_{i_k}) \perp\!\!\!\perp (Y_{j_1},\dots,Y_{j_\ell})$$

Proof. $\boxed{\Rightarrow}$ This follows from the composition principle with the functions
$f:(x_i)_{i\in I}\longmapsto(x_{i_1},\dots,x_{i_k})$ and $g:(y_j)_{j\in J}\longmapsto(y_{j_1},\dots,y_{j_\ell})$

$\boxed{\Leftarrow}$ We apply the previous proposition with

$$\mathcal{C}_1=\bigcup_{\substack{k\geq 1\\ i_1,\dots,i_k\in I}}\sigma(X_{i_1},\dots,X_{i_k}) \quad\text{and}\quad \mathcal{C}_2=\bigcup_{\substack{\ell\geq 1\\ j_1,\dots,j_\ell\in J}}\sigma(Y_{j_1},\dots,Y_{j_\ell}),$$

which are π-systems containing Ω such that $\sigma(\mathcal{C}_1)=\sigma((X_i)_{i\in I})$, $\sigma(\mathcal{C}_2)=\sigma((Y_j)_{j\in J})$
and $\forall A\in\mathcal{C}_1$, $\forall B\in\mathcal{C}_2$, $\mathbb{P}(A\cap B)=\mathbb{P}(A)\,\mathbb{P}(B)$

We now introduce the notion of tail σ-field.

Definition If $(X_i)_{i\geq 1}$ are random variables, we set $\mathcal{B}_n=\sigma(X_k : k\geq n)$ and

$$\mathcal{B}_\infty=\bigcap_{n\geq 1}\mathcal{B}_n,$$

called the tail σ-field of $(X_i)_{i\geq 1}$

Intuitively $\mathcal{B}_\infty$ corresponds to the events which do not change if we change the values of a finite number of X_i's.

Example If $(X_i)_{i\geq 1}$ are real-valued, set $S_n=X_1+\dots+X_n$. Then
$\{\sup_n S_n=+\infty\}\in\mathcal{B}_\infty$.
Indeed, for every $k\geq 1$, this event belongs to $\{\sup_{n\geq k}(X_k+X_{k+1}+\dots+X_n)=+\infty\}\in\mathcal{B}_k$

Theorem (Kolmogorov 0-1 law) Assume that $(X_i)_{i\geq 1}$ are independent
then $\forall B\in\mathcal{B}_\infty$, $\mathbb{P}(B)=0$ or $\mathbb{P}(B)=1$

Proof: Set $\mathcal{D}_n = \sigma(X_1, \dots, X_n)$. Then $\mathcal{D}_n \perp\!\!\!\perp \mathcal{B}_{n+1}$, hence $\mathcal{D}_n \perp\!\!\!\perp \mathcal{B}_\infty$ since $\mathcal{B}_\infty \subset \mathcal{B}_{n+1}$.

Therefore $\forall A \in \bigcup_{n=1}^{\infty} \mathcal{D}_n$, $\forall B \in \mathcal{B}_\infty$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

But $\bigcup_{n=1}^{\infty} \mathcal{D}_n$ is a π-system and $\sigma\left(\bigcup_{n=1}^{\infty} \mathcal{D}_n\right) = \sigma\left((X_n)_{n\geq 1}\right)$

Hence $\forall A \in \sigma(X_n : n \geq 1)$, $\forall B \in \mathcal{B}_\infty$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$

Hence $\forall A \in \mathcal{B}_\infty$, $\forall B \in \mathcal{B}_\infty$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

Taking $A = B$, we get $\mathbb{P}(B) = \mathbb{P}(B)^2$, then $\mathbb{P}(B) = 0$ or 1.

4) Real-valued random variables

Recall that if $(x_n)_{n\geq 1}$ is a sequence in $\mathbb{R}$, we write

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{k\geq n} x_k \in \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$$

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf_{k\geq n} \in \overline{\mathbb{R}}$$

Proposition Let $f_n : (E, \mathcal{E}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ be measurable functions. Then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ are measurable

(We equip $\overline{\mathbb{R}}$ with the distance $d(x,y) = |\mathrm{Arctan}\, x - \mathrm{Arctan}\, y|$

Proof: We show it for $f = \sup_n f_n$ (it is similar for the other function). It is enough to show that $\forall a \in \mathbb{R}$, $f^{-1}([-\infty, a]) \in \mathcal{E}$.

But $f^{-1}([-\infty, a]) = \bigcap_{n\geq 1} f_n^{-1}([-\infty, a]) \in \mathcal{E}$.

Take home message: any "reasonable" operation on measurable functions gives a measurable function

Definition A simple function $f:(E,\mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable function taking a finite number of values. Equivalently, f can be written in the form $f(x) = \sum_{i=1} a_i \mathbb{1}_{x \in A_i}$ $(A_i \in \mathcal{E},\ a_i \in \mathbb{R})$

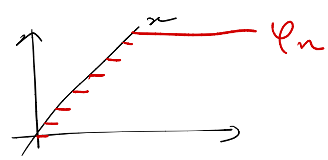
Remark By taking intersections, a simple function can be written uniquely in the form $f(x) = \sum_{i=1} b_i \mathbb{1}_{x \in B_i}$ with $b_1 < b_2 < \cdots < b_n$ and $(B_i)_{1 \leq i \leq m}$ disjoint.

Theorem Let $f: E \to \mathbb{R}^+$ be a measurable function. There exists a sequence (f_n) of simple functions s.t. $0 \leq f_n \uparrow f$, that is $\forall x \in E$, the sequence $(f_n(x))_{n \geq 1}$ is increasing with limit $f(x)$

In practice, this result is useful to show a result on general measurable functions by first showing it for simple functions and then by passing to the limit.

Proof: Step 1: We approximate $\begin{array}{l} \mathbb{R}_+ \longrightarrow \mathbb{R}_+ \\ x \longmapsto x \end{array}$

We just set $\varphi_n(x) = \min(2^{-n} \lfloor 2^n x \rfloor, n)$



Step 2: Take $f_n = \varphi_n \circ f$.

Important application: Doob–Dynkin lemma

Let $f: (E, \mathcal{E}) \longrightarrow (F, \mathcal{F})$ and $g: (E, \sigma(f)) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable functions. Then we can write

$g = h \circ f$ with $h: (F, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a measurable function

Proof: To simplify, assume $g \geq 0$ (otherwise just write $g = g^+ - g^-$ with $g^+ = \max(g, 0) \geq 0$ and $g^- = \max(-g, 0) \geq 0$)

Step 1: Assume that $g = \mathbb{1}_A$ with $A \in \sigma(f)$. Then $A = f^{-1}(B)$ with $B \in \mathcal{F}$.
Setting $h = \mathbb{1}_B$, we have $g(x) = h(f(x))$ because $f(x) \in B \iff x \in f^{-1}(B)$.

Step 2: By linearity, the results holds when g is a simple function.
Now let g_n be $\sigma(f)$-measurable simple functions s.t. $0 \leq g_n \uparrow g$ where $g_n = h_n \circ f$ with $h_n : (F, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable.

Then set
$$h(y) = \begin{cases} \sup_{n \geq 1} h_n(y) & \text{if this quantity is finite} \\ 0 & \text{otherwise} \end{cases}$$

Then for $x \in E$ $\sup_{n \geq 1} h_n(f(x)) = \sup_{n \geq 1} g_n(x) = g(x)$, which implies $g(x) = h(f(x))$.

Remark A simple adaptation of the proof shows that the result is true when g is $\mathbb{R}^n$-valued (but it is false in full generality: see exercise sheet)

Remark In probability, this result is often used to say that $\sigma(X)$ measurable functions are functions of X.

END OF LECTURE 7

5) Integration

The notion of expectation is defined in probability using integration with respect to a measure in measure theory. Since it plays a crucial role, let us recall the setting.

We start with ≥ 0 functions

Let $(E, \mathcal{E}, \mu)$ be a measured space.

a) Definition of the Lebesgue integral

Definition If $f : E \to [0, +\infty]$ is a simple function, $f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ with $a_i \in \mathbb{R}_+ \cup \{+\infty\}$ and $A_i \in \mathcal{E}$, we define $\int f \, d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$ [with the convention $0 \times \infty = 0$]

Remark
- we sometimes write $\int_E f \, d\mu$, $\int f(x) \mu(dx)$, $\int f(x) \, d\mu(x)$
- if $f = \sum_{i=1}^{m} b_i \mathbb{1}_{B_i}$, $\sum_{i=1}^{m} b_i \mu(B_i) = \sum_{i=1}^{n} a_i \mu(A_i)$ (the definition of $\int f d\mu$ does not depend on the choice of the representative of f)

Elementary properties Let $f, g \geq 0$ be simple functions

1) If $a, b \geq 0$, $\int (af + bg)\, d\mu = a\int f\, d\mu + b\int f\, d\mu$

2) If $f \leq g$, $\int f\, d\mu \leq \int g\, d\mu$

In particular, $\int f d\mu = \sup\limits_{\substack{h \text{ simple} \\ h \leq f}} \int h\, d\mu$

Definition Let $f : E \to [0, \infty]$ be measurable. We set

$$\int f\, d\mu = \sup \left\{ \int h\, d\mu : 0 \leq h \leq f,\ h \text{ simple} \right\} \in \mathbb{R}_+ \cup \{+\infty\}$$

In probability, if $X : \Omega \to \mathbb{R}_+$ is a random variable, we <u>define</u>

$$\mathbb{E}[X] := \int_\Omega X(\omega)\, \mathbb{P}(d\omega)$$

In particular, if $A \in \mathcal{F}$, $\mathbb{E}[\mathbb{1}_A] = \int_\Omega \mathbb{1}_A(\omega)\, \mathbb{P}(d\omega) = \mathbb{P}(A)$.

Proposition 1) If $0 \leq f \leq g \leq +\infty$, $\int f\, d\mu \leq \int g\, d\mu$

2) If $f \geq 0$ and $\mu(\{x \in E : f(x) > 0\}) = 0$, then $\int f\, d\mu = 0$.

Proof: 1) Ok by definition

2) Let h be a simple function such that $0 \leq h \leq f$.

Since $h(x) > 0 \Rightarrow f(x) > 0$, we have $\mu(\{x \in E : h(x) > 0\}) = 0$.

Hence $\int h\, d\mu = 0$ by definition of the integral of a ≥ 0 simple function

So $\int f\, d\mu = 0$ by definition of the integral of a ≥ 0 function

b) Monotone convergence

Theorem (monotone convergence)

Let $(f_n : E \to [0, +\infty])_{n \geq 1}$ be a non-decreasing sequence of functions

(i.e. $\forall x \in E$, $\forall n \geq 1$, $f_{n+1}(x) \geq f_n(x)$). Set $f(x) = \lim\limits_{n \to \infty}^{\uparrow} f_n(x)$

means that the sequence is increasing

Then $\int f\, d\mu = \lim\limits_{n \to \infty}^{\uparrow} \int f_n\, d\mu \in [0, \infty]$

Probabilistic version: If $(X_n)_{n\geq 0}$ is a non-decreasing sequence of random variables, $\mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$

Here $\lim_{n\to\infty} X_n$ is the random variable $\omega \mapsto \lim_{n\to\infty} X_n(\omega)$

Proof ① Since $f \geq f_n$, we have $\int f d\mu \geq \int f_n d\mu$. Hence $\int f d\mu \geq \lim_{n\to\infty} \int f_n d\mu$
(The limit exists since $(\int f_n d\mu)_{n\geq 1}$ is increasing)

② For the other inequality, take h simple, $h \leq f$
The goal is to connect h to f_n.

Idea 1: Set $E_n = \{x \in E : h(x) \leq f_n(x)\}$
problem: it can be empty for every $n \geq 1$.

Idea 2: introduce a new parameter $a > 1$ to "get some space": set
$E_n = \{x : h(x) \leq a f_n(x)\}$.

Since $h(x) \leq a f(x)$ $\forall x \in E$, we have $\bigcup_{n\geq 1} E_n = E$, and $f_n \geq \frac{1}{a} \mathbb{1}_{E_n} h$.

Hence $\int f_n d\mu \geq \int \frac{1}{a} \mathbb{1}_{E_n} h \, d\mu = \frac{1}{a} \sum_{i=1}^{k} b_i \mu(B_i \cap E_n)$ if $h = \sum_{i=1}^{k} b_i \mathbb{1}_{B_i}$

But $B_i \cap E_n \uparrow B_i$

Hence $\lim_{n\to\infty} \mu(B_i \cap E_n) = \mu(B_i)$

Therefore $\liminf_{n\to\infty} \int f_n d\mu \geq \frac{1}{a} \sum_{i=1}^{k} b_i \mu(B_i) = \frac{1}{a} \int h \, d\mu$

Take $a \to 1$: $\liminf_{n\to\infty} \int f_n d\mu \geq \int h \, d\mu$.

Taking the sup over h simple, we get $\lim_{n\to\infty} \int f_n d\mu \geq \int f d\mu$

Corollary ① If $f, g \geq 0$ and $a, b \geq 0$, $\int (af + bg) d\mu = a \int f d\mu + b \int g \, d\mu$

② If $f_k \geq 0$, $\int (\sum f_k) d\mu = \sum_k \int f_k d\mu$

Probabilistic version: • for $X, Y \geq 0$ r.v. $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
• for $X_k \geq 0$ r.v., $\mathbb{E}[\sum_k X_k] = \sum_k \mathbb{E}[X_k]$

Proof: ① Let f_n, g_n be simple ≥ 0 functions such that $f_n \uparrow f$, $g_n \uparrow g$ (cf 4) for existence). Then by monotone convergence:

$$\int (af+bg)\,d\mu = \lim \uparrow \int (a f_n + b g_n)\,d\mu$$
$$= \lim \uparrow \left(a\int f_n\,d\mu + b\int g_n\,d\mu\right)$$
$$= a\int f\,d\mu + b\int g_n\,d\mu$$

② Set $F_n = \sum_{k=1}^{n} f_k$, $F = \sum_{k=1}^{\infty} f_k$. Then $F_n \uparrow F$, so

$$\int F_n\,d\mu \to \int F\,d\mu$$

But $\int F_n\,d\mu = \sum_{k=1}^{n} \int f_k\,d\mu \xrightarrow[n\to\infty]{} \sum_{k=1}^{\infty} \int f_k\,d\mu$

$\sim$

Examples. For $a \in E$, the Dirac measure at a is defined by $\delta_a(A) = \begin{cases} 0 & \text{if } a \notin A \\ 1 & \text{if } a \in A \end{cases}$ for every $A \in \mathcal{E}$. Then $\forall f: E \to \mathbb{R}_+$ measurable,

$$\int_E f(x)\,\delta_a(dx) = f(a).$$

Indeed, this is true for simple functions and we conclude by monotone convergence

- If $\#$ is the counting measure on $\mathbb{N}$, then $\forall f: \mathbb{N} \to \mathbb{R}_+$,

$$\int_{\mathbb{N}} f(x)\,\#(dx) = \sum_{i=0}^{\infty} f(i)$$

Indeed, this is true for simple functions and we conclude by monotone convergence

c) Fatou's lemma

Theorem (Fatou's lemma) Let (f_n) be a sequence of ≥ 0 measurable functions. Then

$$\int (\liminf_{n\to\infty} f_n)\, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu$$

Probabilistic setting: If $(X_k)_{k\geq 1}$ are ≥ 0 r.v, $\mathbb{E}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]$

Proof: By definition, $\liminf_{n\to\infty} f_n = \lim_{k\to\infty}\uparrow (\inf_{n\geq k} f_k)$, so by monotone convergence

$\int (\liminf_{n\to\infty} f_n)\, d\mu = \lim_{k\to\infty}\uparrow \int (\inf_{n\geq k} f_n)\, d\mu$. But for every integer $p \geq k$, $\inf_{n\geq k} f_n \leq f_p$, which implies $\int (\inf_{n\geq k} f_n)\, d\mu \leq \inf_{p\geq k} \int f_p \, d\mu$. By taking the increasing limit as $k \uparrow \infty$, we get the result

$\sim$

d) Markov's inequality

We say that a property is true almost everywhere (a.e) if the set of $x \in E$ for which it is not true is negligeable, i.e. has 0 measure (in probability: almost-surely / a.s)

Proposition Let $f \geq 0$

1. $\forall a > 0$, $\mu(\{x : f(x) \geq a\}) \leq \frac{1}{a} \int f \, d\mu$
2. $\int f \, d\mu < \infty \Rightarrow f < \infty$ a.e
3. $\int f \, d\mu = 0 \Rightarrow f = 0$ a.e
4. If $g \geq 0$, $f = g$ a.e $\Rightarrow \int f \, d\mu = \int g \, d\mu$

Probabilistic setting for $X, Y \geq 0$ r.v.

1. $\forall a > 0$, $\mathbb{P}(X \geq a) \leq \frac{1}{a} \mathbb{E}[X]$ (Markov's inequality)
2. $\mathbb{E}[X] < \infty \Rightarrow X < \infty$ a.s.
3. $\mathbb{E}[X] = 0 \Rightarrow X = 0$ a.s
4. $X = Y$ a.s $\Rightarrow \mathbb{E}[X] = \mathbb{E}[Y]$

Proof ① Set $E_a = \{x : f(x) \geq a\}$. Then $f \geq a \mathbb{1}_{E_a}$.

Hence $\int f d\mu \geq \int a \mathbb{1}_{A_a} d\mu = a\mu(A_a)$

② Set $A_n = \{f \geq n\}$ (i.e $A_n = \{x : f(x) \geq n\}$)

and $A_\infty = \{f = \infty\}$.

Then $\mu(A_\infty) = \mu(\bigcap_{n \geq 1} A_n) = \lim_{n\to\infty} \mu(A_n) \overset{①}{\leq} \lim_{n\to\infty} \frac{1}{n} \int f d\mu = 0$.

③ $\boxed{\Leftarrow}$ Already seen

$\boxed{\Rightarrow}$ Set $B_n = \{f \geq \frac{1}{n}\}$. Then $\mu(B_n) \leq n \int f d\mu = 0$

Hence $\mu(\{f > 0\}) = \mu(\bigcup_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} \mu(B_n) = 0$

④ Write $\int \max(f,g) d\mu = \int \min(f,g) d\mu + \int (\underbrace{\max(f,g) - \min(f,g)}_{=0 \text{ a.e.}}) d\mu = \int \min(f,g) d\mu$.

But $\min(f,g) \leq f, g \leq \max(f,g)$

Hence $\int \min(f,g) d\mu = \int f d\mu = \int g d\mu = \int \max(f,g) d\mu$

e) Fubini's Theorem

Recall that μ is σ-finite if one can write $E = \bigcup_{n=1}^{\infty} A_n$ with $(A_n)_{n\geq 1}$ countable sequence of events with $\mu(E_n) < \infty$ $\forall n \geq 1$.

Informally, Fubini's theorem states that if $f : \mathbb{R}^n \to \mathbb{R}_+$ is measurable, $\mu_1, \dots, \mu_n$ are σ-finite, then the integral $\int\int\cdots\int f(x_1, \dots, x_n) \mu_1(dx_1) \cdots \mu_n(dx_n)$ can be computed by integrating in any order.

In probability, this means that one can exchange $\mathbb{E}[\cdot]$ and $\int dx$ for ≥ 0 r.v.

We state the theorem for $n=2$

Theorem (Fubini-Tonelli) Let μ, ν be σ-finite measures on $(E, \mathcal{E})$ and $(F, \mathcal{F})$. We equip $E \times F$ with the product σ-field $\mathcal{E} \otimes \mathcal{F}$. Let $f : E \times F \to \mathbb{R}_+$ be measurable.

(1) $x \mapsto \int f(x,y) \nu(dy)$ and $y \mapsto \int f(x,y) \mu(dx)$ are measurable

(2) We have $\int_{E\times F} f \, d\underset{\uparrow \text{ product measure}}{\mu\otimes\nu} = \int_E \left(\int_F f(x,y) \nu(dy) \right) \mu(dx) = \int_F \left(\int f(x,y) \mu(dx) \right) \nu(dy)$.

We do not give the proof here: it involves additional inputs from measure theory

8) Integrating real-valued random variables

If $f : E \to \mathbb{R}$ is real valued, when $\int |f| \, d\mu < \infty$, we say that f is μ-integrable and write $f \in L^1(E, \mathcal{E}, \mu)$

More generally for $p>0$, when $\int |f|^p d\mu < \infty$ we write $f \in L^p(E, \mathcal{E}, \mu)$

Definition When $\int |f| \, d\mu < \infty$, write $f = f_+ - f_-$ with $f_+ = \max(f, 0) \geq 0$ and $f_- = \max(-f, 0) \geq 0$ and define $\int f \, d\mu = \int f_+ \, d\mu - \int f_- \, d\mu$.

As for ≥ 0 functions one checks that for f, g μ-integrable:

- $|\int f d\mu| \leq \int |f| d\mu$ (triangular inequality)
- $f \leq g$ μ almost everywhere implies $\int f \, d\mu \leq \int g \, d\mu$
- $f = g$ μ almost everywhere implies $\int f d\mu = \int g d\mu$.

Probabilistic setting $X : \Omega \to \mathbb{R}$ is integrable (in $L^1(\Omega, \mathcal{A}, \mathbb{P})$) if $\mathbb{E}[|X|] < \infty$.

When $\mathbb{E}[|X|^p] < \infty$ we write $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$

(observe that if X is bounded, i.e $|X| \leq C$ for a fixed (deterministic) C, then $\mathbb{E}[|X|^p] \leq C^p$ so $X \in L^p$)

For real-valued functions, we have the following very important result.

Theorem (dominated convergence) Let (f_n) be a sequence of integrable real-valued measurable functions. Assume

① $\exists f : E \to \mathbb{R}$ measurable such that $f_n(x) \xrightarrow[n \to \infty]{} f(x)$ holds for μ-almost all x

② $\exists g : E \to \mathbb{R}_+$ measurable such that $\int g d\mu < \infty$ and $\forall n \geq 1$, $|f_n(x)| \leq g(x)$ for μ-almost all x [domination condition]

Then $\int |f_n - f| d\mu \xrightarrow[n \to \infty]{} 0$ and $\int f_n d\mu \to \int f d\mu$

Probabilistic setting Let X_n be real-valued r.v. Assume that:

① $\exists X$ r.v. such that $X_n \xrightarrow[n \to \infty]{} X$ almost surely

② $\exists Z \geq 0$ r.v. such that $\mathbb{E}[Z] < \infty$ and $\forall n \geq 1$, $|X_n| \leq Z$ almost surely

Then $\mathbb{E}[|X_n - X|] \xrightarrow[n \to \infty]{} 0$, $\mathbb{E}[X_n] \xrightarrow[n \to \infty]{} \mathbb{E}[X]$

Remark In the domination condition, it is crucial that the dominating function must NOT depend on n.

Proof: By (1) and (2), $|f| \leq g$ μ almost everywhere, so $\int |f| d\mu \leq \int g \, d\mu < \infty$, which shows that f is integrable.

Next, the idea is to consider $h_n = 2g - |f - f_n|$.

Observe that $h_n \geq 0$ μ-almost every where, so that $h_n = h_n \mathbb{1}_{h_n \geq 0}$ μ almost every where.

By Fatou's lemma: $\int \liminf_{n\to\infty} (h_n \mathbb{1}_{h_n \geq 0}) d\mu \leq \liminf_{n\to\infty} \int h_n \mathbb{1}_{h_n \geq 0} d\mu = \liminf_{n\to\infty} \int h_n \, d\mu$

$\int \liminf_{n\to\infty} (h_n \mathbb{1}_{h_n \geq 0}) d\mu = \int 2g \, d\mu$.

Thus $2\int g \, d\mu \leq 2g - \limsup_{n\to\infty} \int |f - f_n| \, d\mu$, which implies $\int |f - f_n| d\mu \to 0$.

The second convergence follows from $\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| d\mu$.

∞

The extension of Fubini–Tonelli to real-valued functions is the following

Theorem (Fubini–Lebesgue) Let μ, ν be σ-finite measures on $(E, \mathcal{E})$ and $(F, \mathcal{F})$. We equip $E \times F$ with the product σ-field $\mathcal{E} \otimes \mathcal{F}$. Let $f : E \times F \to \mathbb{R}$ be $\mu \otimes \nu$ integrable. Then

① for μ-almost every x, $y \mapsto f(x,y)$ is ν-integrable and for ν almost every y, $x \mapsto f(x,y)$ is μ-integrable

② The functions $x \mapsto \int f(x,y) \, \nu(dy)$ and $y \mapsto \int f(x,y) \, \mu(dx)$ are well defined, except maybe on sets with 0 measure, and are respectively μ and ν integrable.

③ We have $\int_{E \times F} f \, d\mu \otimes \nu = \int_E \left(\int_F f(x,y) \nu(dy) \right) \mu(dx) = \int_F \left(\int_E f(x,y) \, \mu(dx) \right) \nu(dy)$

In practice, one first checks if $|f|$ is $\mu \otimes \nu$ integrable using Fubini–Tonelli, and then one can interchange the order of integration.

Application If X is an integrable real-valued r.v. and $(A_i)_{i \geq 1}$ are events such that $\Omega = \bigcup_{i \geq 1} A_i$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{E}[X \mathbb{1}_{A_i}]$$

Indeed, $\sum_{i=1}^{\infty} \mathbb{E}[X \mathbb{1}_{A_i}] = \mathbb{E}[X \underbrace{(\sum_{i=1}^{\infty} \mathbb{1}_{A_i})}_{1}]$

1 because $\forall \omega \in \Omega$ there is only one $i \geq 1$ such that $\omega \in A_i$

This is useful when one to treat several different cases.

It can be seen as a functional version of the law of total probability, which is:

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap A_i) \quad \left(\text{which comes from the fact that } A = \bigcup_{i=1}^{\infty} A \cap A_i\right)$$

Example: If Y is a $\mathbb{Z}$-valued r.v and $X \geq 0$, $\mathbb{E}[X] = \sum_{i \in \mathbb{Z}} \mathbb{E}[X \mathbb{1}_{Y=i}]$

6) Classical laws

a) Discrete laws

- uniform law: If E is a finite set with n elements, X follows the uniform distribution on E if $\mathbb{P}(X = x) = \frac{1}{n}$ for all $x \in E$

- Bernoulli law of parameter $p \in [0,1]$: it's the law of a r.v X in $\{0,1\}$ with $\mathbb{P}(X=1)=p$ $\mathbb{P}(X=0)=1-p$
Interpretation: result of a rigged coin giving heads with probability p

- Binomial law $B(n,p)$ $(n \geq 1, p \in [0,1])$ it's the law of a r.v. in $\{1, \dots, n\}$ with $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$
Interpretation: number of heads when tossing n times the previous coin

- Geometric law of parameter $p \in (0,1)$. It's the law of a r.v in $\{0,1,2,\dots\}$ with $\mathbb{P}(X=k) = p(1-p)^{k-1}$ $k \geq 1$.
Interpretation: number of trials before the first heads with the previous coin

- Poisson law of parameter $\lambda > 0$: it's the law of a r.v. in $\mathbb{N}$ with $\mathbb{P}(X=k) = \frac{\lambda^k}{k!} e^{-k}$, $k \geq 0$
Interpretation: law of rare events (this will be seen later)

b) Continuous laws

END OF LECTURE 8

This involves the notion of density

Definition Let $p : \mathbb{R} \to \mathbb{R}_+$ be a measurable function such that $\int_{\mathbb{R}} p(x)\,dx = 1$
For $A \in \mathcal{B}(\mathbb{R})$, the formula $\mu(A) = \int_A p(x)\,dx = \int \mathbb{1}_A(x)\, p(x)\,dx$ defines a probability measure on $\mathbb{R}$
A r.v. X having this as law is called a r.v. with density p. It's cdf is $\mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(t)\,dt$.

⚠ a density is not uniquely defined: it is defined up to a 0 Lebesgue measure set (if $p = q$ almost everywhere, X has density p and q ⚠

One can check that for every $f : \mathbb{R} \to \mathbb{R}_+$ measurable $\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\, p(x)\, dx$ $(\ast)$

Indeed, if f is the form $f = \mathbb{1}_A$ with $A \in \mathcal{B}(\mathbb{R})$, we have

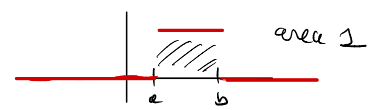
$$\mathbb{E}[f(X)] = \int_{\Omega} \mathbb{1}_A(X(\omega))\, \mathbb{P}(d\omega) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}_X(A)$$ (definition of the integral of a simple function)

Also $\int_{\mathbb{R}} \mathbb{1}_A(x)\, p(x)\, dx = \int_A p(x)\, dx = \mathbb{P}_X(A)$ by definition.

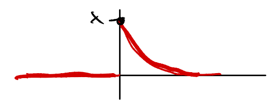
By linearity $(\ast)$ holds for f simple. Then take $0 \leq f_n \nearrow f$ with f_n simple.

Then $\underbrace{\mathbb{E}[f_n(X)]}_{\xrightarrow{n\to\infty} \mathbb{E}[f(X)]} = \int_{\mathbb{R}} f_n(x) p(x)\, dx \xrightarrow{n\to\infty} \int f(x) p(x)\, dx$ by monotone convergence (twice).

- Uniform law on $[a,b]$ $(a<b)$: $\quad p(x) = \dfrac{1}{b-a} \mathbb{1}_{[a,b]}(x)$



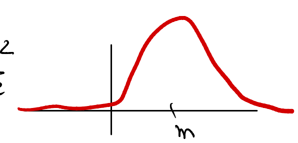
- Exponential law of parameter $\lambda > 0$ $\quad p(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}$



Then $\mathbb{E}[X] = \int_0^\infty \lambda x e^{-\lambda x}\, dx = \dfrac{1}{\lambda}$

It satisfies $\mathbb{P}(X > a+b) = \mathbb{P}(X > a)\, \mathbb{P}(X > b)$ (memoryless property)

- Gaussian law $N(m, \sigma^2)$ $(m \in \mathbb{R}, \sigma^2 > 0)$



One checks that $\mathbb{E}[X] = m$, $\sigma^2 = \mathbb{E}[(X-m)^2]$ $\quad p(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$

Remark If X has density p, its CDF $F(x) = \int_{-\infty}^{x} p(u)\, du$ is continuous. Indeed, take a sequence $x_n \to x$.

Then $F(x_n) = \int_{\mathbb{R}} \underbrace{p(u)\, \mathbb{1}_{]-\infty, x_n]}(u)}_{f_n(u)}\, du$.

- For fixed $u \neq x$, $f_n(u) \longrightarrow p(u)\, \mathbb{1}_{]-\infty, x]}(u)$ since $\{x\}$ has 0 Lebesgue measure, this convergence holds for almost all u
- $|f_n(u)| \leq p(u)$, which is an integrable function which does not depend on n

Thus $F(x_n) \longrightarrow \int_{\mathbb{R}} p(u)\, \mathbb{1}_{]-\infty, x]}(u)\, du = F(x)$

⚠ The converse is false in general: CDF continuous $\not\Rightarrow$ $\exists$ density

What is true is CDF of X continuous $\Leftrightarrow$ $\forall x \in \mathbb{R}$, $\mathbb{P}(X = x) = 0$.

Remark If $F: \mathbb{R} \to [0,1]$ is non-decreasing, $\lim_{-\infty} F = 0$, $\lim_{+\infty} F = 1$, continuous AND piece-wise C^1, that is $\exists -\infty = a_0 < a_1 < \dots < a_n = \infty$ such that F is C^1 on (a_i, a_{i+1}) for every $0 \leq i \leq n-1$, then F is the CDF of a r.v with density $p(x) = \begin{cases} F'(x) & \text{for } x \notin \{a_1, \dots, a_{n-1}\} \\ 0 & \text{otherwise} \end{cases}$

Indeed, for every $x \in \mathbb{R}$ we then have $F(x) = \int_{-\infty}^{x} p(t)\,dt$

In particular, if the cdf of X is continuous and piece-wise C^1, then X has a density (given by the a.s. derivative of its cdf)

Remark One similarly defines random variables with density on $\mathbb{R}^n$: if $p: \mathbb{R}^n \to \mathbb{R}_+$ is a measurable function with $\int_{\mathbb{R}^n} p(x_1, \dots, x_n)\,dx_1 \dots dx_n$, $X = (X_1, \dots, X_n)$ has density p if $\forall A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(X \in A) = \int_A p(x_1, \dots, x_n)\,dx_1 \dots dx_n.$$

Then $\forall\, 1 \leq i \leq n$ X_i has a density p_i in $\mathbb{R}$ obtained by integrating p with respect to the other variables:

$$p_i(x) = \int_{\mathbb{R}^{n-1}} p(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)\,dx_1 \dots dx_{i-1}\,dx_{i+1} \dots dx_n.$$

7) Integration and independence

The following result is very useful

Theorem (transfer theorem) Let $X: \Omega \to E$ be a random variable, $f: E \to \mathbb{R}_+$ a measurable function. Then $\mathbb{E}[f(X)] = \int_E f(x)\,\mathbb{P}_X(dx)$

Proof: Step 1: Take $f = \mathbb{1}_A$ with $A \in \mathcal{E}$. Then $\mathbb{E}[\mathbb{1}_{X \in A}] = \mathbb{P}(X \in A)$, and

$$\int_E \mathbb{1}_A(x)\,\mathbb{P}_X(dx) = \mathbb{P}_X(A) = \mathbb{P}(X \in A)$$

Step 2: By linearity, the result is true for any ≥ 0 simple function.

We then take a sequence (f_n) of simple functions such that $0 \leq f_n \leq f$ and $f_n \uparrow f$, and conclude by monotone convergence:

$$\begin{aligned} \mathbb{E}[f_n(X)] &= \int_E f_n(x)\,\mathbb{P}_X(dx) \xrightarrow[n\to\infty]{} \int_E f(x)\,\mathbb{P}_X(dx) \\ \| \quad & \\ \int_\Omega f_n(X(\omega))\,\mathbb{P}(d\omega) &\xrightarrow[n\to\infty]{} \int_\Omega f(X(\omega))\,\mathbb{P}(d\omega) = \mathbb{E}[f(X)]. \end{aligned}$$

Hence $\mathbb{E}[f(X)] = \int_E f(x)\,\mathbb{P}_X(dx)$

~

Remark If X has density p, we have seen that for $f: \mathbb{R} \to \mathbb{R}_+$ measurable, $\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\,p(x)\,dx$.

It is also $\int_{\mathbb{R}} f(x)\,\mathbb{P}_X(dx)$ by the transfer theorem. For this reason we write $\boxed{\mathbb{P}_X(dx) = p(x)\,dx}$

Remark The transfer theorem is also valid for $f : E \to \mathbb{R}$ (not necessarily ≥ 0) bounded (write $f = f^+ - f^-$ with $f^+, f^- \geq 0$) and more generally for $f : E \to \mathbb{R}$ such that $\mathbb{E}[|f(X)|] < \infty$.

Application: Let U be uniform on $[0,1]$. Find the law of U^2.
Take $f : [0,1] \to \mathbb{R}_+$ measurable. Then by the transfer theorem

$$\mathbb{E}[f(U^2)] = \int_{\mathbb{R}} f(u^2)\, \mathbb{P}_U(du) = \int_0^1 f(u^2)\, du \overset{[u^2 = x]}{=} \int_0^1 f(x) \frac{1}{2\sqrt{x}}\, dx$$

But we know that $\mathbb{E}[f(U^2)] = \int_{\mathbb{R}} f(x)\, \mathbb{P}_{U^2}(dx)$

We conclude that $\boxed{\mathbb{P}_{U^2}(dx) = \mathbb{1}_{[0,1]}(x) \frac{1}{2\sqrt{x}}\, dx}$

Indeed, if $\int_E f(x)\, \mu(dx) = \int_E f(x)\, \nu(dx)$ for every $f : E \to \mathbb{R}_+$ measurable, then $\mu = \nu$ (just take $f = \mathbb{1}_A$).
This is the <u>dummy function method</u>.

Application If X has density p, then $\mathbb{E}[|X|] = \int_0^\infty |x|\, \mathbb{P}_X(dx) = \int_0^\infty |x|\, p(x)\, dx$

Example If X has density $\frac{\alpha+1}{x^\alpha} \mathbb{1}_{[1,\infty[}(x)$, then for $p > 0$ $\mathbb{E}[X^p] < \infty \Leftrightarrow p > \alpha + 1$.

Indeed, by the transfer theorem, $\mathbb{E}[X^p] = \int_0^\infty x^p\, \mathbb{P}_X(dx) = \int_1^{+\infty} \frac{\alpha+1}{x^{p-\alpha}}\, dx$ which is finite iff $p - \alpha > 1$.

Corollary If $X : \Omega \to E$ and $Y : \Omega \to E$ are two random variables with same law, then $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$ $\forall f : E \to \mathbb{R}_+$ measurable

<u>proof</u>: $\mathbb{E}[f(X)] = \int_E f(x)\, \mathbb{P}_X(dx) = \int_E f(x)\, \mathbb{P}_Y(dx) = \mathbb{E}[f(Y)]$

Example if $X : \Omega \to \mathbb{R}_+$ and $Y : \Omega \to \mathbb{R}_+$ have the same law, then $\mathbb{E}[X^p] = \mathbb{E}[Y^p]$ $\forall p > 0$.

END OF LECTURE 9

Theorem If $X_1, \dots, X_n$ are ⫫ real-valued random variables with X_i having density p_i, then $(X_1, \dots, X_n)$ has density $p_1(x_1) \cdots p_n(x_n)$

Proof: Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be measurable. We use the dummy function method.

By ⫫, $\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$. Thus

$$\mathbb{E}[f(X_1,\dots,X_n)] \overset{\text{transfer theorem}}{=} \int_{\mathbb{R}^n} f(x_1,\dots,x_n)\, \mathbb{P}_{(X_1,\dots,X_n)}(dx_1 \cdots dx_n) = \int_{\mathbb{R}^n} f(x_1,\dots,x_n)\, \mathbb{P}_{X_1}(dx_1) \otimes \cdots \otimes \mathbb{P}_{X_n}(dx_n)$$

$$= \int_{\mathbb{R}^n} f(x_1,\dots,x_n) \underbrace{p_1(x_1) \cdots p_n(x_n)\, dx_1 \cdots dx_n}_{\text{law of } (X_1,\dots,X_n)}.$$

⚠ In general, if $X_1, \dots, X_n$ have densities in $\mathbb{R}$, $(X_1,\dots,X_n)$ does not have a density in $\mathbb{R}^n$ (if X has a density, one can check that (X,X) does not have a density in $\mathbb{R}^2$).

Theorem Let X, Y be ⫫ r.v. in $\mathbb{R}$ with densities. Then $X+Y$ has a density

Proof Let p, q be densities of X, Y. We use the dummy function method. Let $f : \mathbb{R} \to \mathbb{R}_+$ be measurable.

$$\mathbb{E}[f(X+Y)] = \int_{\mathbb{R}^2} f(x+y)\, p(x) q(y)\, dx\, dy = \int_{\mathbb{R}} q(y) \left(\int_{\mathbb{R}} f(x+y) p(x)\, dx \right) dy$$

To compute $\int_{\mathbb{R}} f(x+y) p(x) dx$ (for fixed y) we use the change of variables $z = x+y$:

$\int_{\mathbb{R}} f(x+y) p(x) dx = \int_{\mathbb{R}} f(z)\, p(z-y)\, dz$, so $\mathbb{E}[f(X+Y)] = \int_{\mathbb{R}} q(y) \left(\int_{\mathbb{R}} f(z)\, p(z-y)\, dz \right) dy$

and using Fubini–Tonelli this is $\int_{\mathbb{R}} f(z) \left(\int_{\mathbb{R}} p(z-y) q(y)\, dy \right) dz$

We conclude that $X+Y$ has a density, given by $z \mapsto \int_{\mathbb{R}} p(z-y) q(y) dy$ (called the convolution of X and Y)

Application Let (X,Y) be a r.v with density in $\mathbb{R}^2$. Then $\mathbb{P}(X=Y) = 0$

Proof Let p be a density of (X,Y). Using the transfer theorem, write

$$\mathbb{P}(X=Y) = \mathbb{E}[\mathbb{1}_{X=Y}] = \int_{\mathbb{R}^2} \mathbb{1}_{x=y}\, p(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{x=y}\, p(x,y)\, dx \right) dy$$

But for fixed $y \in \mathbb{R}$, $x \mapsto \mathbb{1}_{x=y}\, p(x,y)$ is $= 0$ if $x \neq y$.

Thus $x \mapsto 1_{x=y}\, p(x,y)$ is almost everywhere equal to 0. Thus $\int 1_{x=y}\, p(x,y)\, dx = 0$.
Thus $\mathbb{P}(X=Y) = \int_{\mathbb{R}} 0\, dy = 0$

~

Application If X has a density, (X,X) does not have a density (since $\mathbb{P}(X=X)=1$)

Corollary If X, Y are $\perp\!\!\!\perp$ r.v with densities, then $\mathbb{P}(X=Y)=0$
Indeed, by what we have seen (X,Y) has then a density in $\mathbb{R}^2$.

Theorem The following assertions are equivalent:
① $X_1, \dots, X_n$ are $\perp\!\!\!\perp$
② $\forall f_i : E \to \mathbb{R}_+$ measurable, $\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] = \prod_{k=1}^{n} \mathbb{E}[f_k(X_k)]$.

Proof: For ① ⇒ ②, we use the fact that $X_1, \dots, X_n \perp\!\!\!\perp \Rightarrow \mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$. By the transfer theorem

$$\begin{aligned}
\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] &= \int f_1(x_1) \cdots f_n(x_n)\, \mathbb{P}_{(X_1,\dots,X_n)}(dx_1 \cdots dx_n) \\
&= \int f_1(x_1) \cdots f_n(x_n)\, \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}(dx_1 \cdots dx_n) \\
&\overset{[\text{Fubini}]}{=} \prod_{k=1}^{n} \int f_k(x_k)\, \mathbb{P}_{X_k}(dx_k) \\
&= \prod_{k=1}^{n} \mathbb{E}[f_k(X_k)]
\end{aligned}$$

For ② ⇒ ① just take $f_i = 1_{A_i}$ for A_i measurable to get
$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n)$

~

In practice, to show that $X \perp\!\!\!\perp Y$, one often computes $\mathbb{E}[f(X)g(Y)]$ to obtain something of the form $\left(\int f(x)\mu(dx)\right)\left(\int g(y)\,\nu(dx)\right)$.

Corollary if $(X_1, \dots, X_n)$ has a density of a product form $g_1(x_1) \cdots g_n(x_n)$ then $X_1, \dots, X_n$ are $\perp\!\!\!\perp$

This readily follows from ② ⇒ ① in the previous Theorem

Remark If the functions f_i are real valued, the equality $\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] \overset{(\#)}{=} \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$ is true under the integrability condition $\mathbb{E}[|f_i(X_i)|] < \infty$ for every $1 \leq i \leq n$.
In particular, if $X_1, \dots, X_n$ are $\perp\!\!\!\perp$ integrable r.v, then $X_1 \times X_2 \times \dots \times X_n$ is integrable and $\mathbb{E}\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n \mathbb{E}[X_i]$

WARNING in general this is false without the $\perp\!\!\!\perp$ condition (take $n=2$, $X_1 = X_2$ with $\mathbb{E}[X_1]=0$)

Application ① Let X be a real-valued in L^2. Then $X \in L^1$ and we define the variance of X by $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$

② [Cauchy-Schwarz] let X be in L^2 then $\mathbb{E}[|X|]^2 \leq \mathbb{E}[X^2]$

③ Let $(X_i)_{1 \leq i \leq n}$ be independent real-valued L^2 random variables. Then

$$\mathrm{Var}(X_1 + \dots + X_n) = \mathrm{Var}(X_1) + \dots + \mathrm{Var}(X_n)$$

Informally, the variance of a random variable measures its "dispersion" from its mean.

Proof ① Write

$$\begin{aligned}
\mathbb{E}[|X|] &= \mathbb{E}[|X| \mathbb{1}_{|X| \leq 1}] + \mathbb{E}[|X| \mathbb{1}_{|X|>1}] \\
&\leq \mathbb{E}[1 \cdot \mathbb{1}_{|X| \leq 1}] + \mathbb{E}[X^2 \mathbb{1}_{|X|>1}] \\
&\leq 1 + \mathbb{E}[X^2] < \infty.
\end{aligned}$$

To prove the formula for $\mathrm{Var}(X)$, write using linearity of expectation:

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

② This follows from $0 \leq \mathrm{Var}(|X|) = \mathbb{E}[|X|^2] - \mathbb{E}[|X|]^2 = \mathbb{E}[X^2] - \mathbb{E}[|X|]^2$

③ It suffices to establish the result for $n=2$ (the general case follows by induction using the fact that $X_1 \perp\!\!\!\perp X_2 + \dots + X_n$).

Write

$$\begin{aligned}
\mathrm{Var}(X_1+X_2) &= \mathbb{E}[(X_1+X_2)^2] - (\mathbb{E}[X_1+X_2])^2 \\
&= \mathbb{E}[X_1^2] + 2\mathbb{E}[X_1 X_2] + \mathbb{E}[X_2^2] - (\mathbb{E}[X_1]^2 + 2\mathbb{E}[X_1]\mathbb{E}[X_2] + \mathbb{E}[X_2]^2)
\end{aligned}$$

But $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$ by $\perp\!\!\!\perp$.

Thus

$$\begin{aligned}
\mathrm{Var}(X_1 + X_2) &= (\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) + (\mathbb{E}[X_2^2] - \mathbb{E}[X_2]^2) \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2)
\end{aligned}$$

Application If X is a $\mathcal{B}(n,p)$ random variable, then $X \overset{law}{=} Y_1 + \cdots + Y_n$ with $(Y_i)_{1 \le i \le n}$ independent Bernoulli(p) random variables, so $Var(X) = Var(Y_1 + \cdots + Y_n) = Var(Y_1) + \cdots + Var(Y_n) = p(1-p) + \cdots + p(1-p)$

$$= n\,p(1-p).$$

Remarks. • We have $Var(aX+b) = a^2\,Var(X)$

• ③ above is false in general without $\perp\!\!\!\perp$ (take $n=2$ $X_1 = X_2$)