1 Review of Probability

1.1 Events and Probability

Definition 1.1

Let Ω be a non-empty set. A *σ-field $\mathcal{F}$* on Ω is a family of subsets of Ω such that

1) the empty set $\emptyset$ belongs to $\mathcal{F}$;
2) if A belongs to $\mathcal{F}$, then so does the complement $\Omega \setminus A$;
3) if $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$, then their union $A_1 \cup A_2 \cup \cdots$ also belongs to $\mathcal{F}$.

A σ-field on a non-empty set is a family of subsets of the non-empty set.

* 区分 Borel Set 和 σ-field

Example 1.1

Throughout this course $\mathbb{R}$ will denote the set of real numbers. The family of *Borel sets $\mathcal{F} = \mathcal{B}(\mathbb{R})$* is a σ-field on $\mathbb{R}$. We recall that $\mathcal{B}(\mathbb{R})$ is the smallest σ-field containing all intervals in $\mathbb{R}$.

$\mathcal{B}(\mathbb{R}) \subset 2^{\mathbb{R}}$ (power set of $\mathbb{R}$)

Example 1.1. Let $\mathbb{R}$ be the set of real numbers. Denote $\mathcal{B}(\mathbb{R})$ (or, $\mathcal{B}$) the smallest σ-field containing all intervals in $\mathbb{R}$, i.e.,

$$[a,b],\ [a,b),\ (a,b),\ (a,\infty) \in \mathcal{B},$$

for any $a, b \in \mathbb{R}$. Then, it is called the Borel[4] σ-field, and a set $A \in \mathcal{B}$ is called *a Borel set.* □

Remark. From the definitions, we have that, for any real number $a \in \mathbb{R}$ and any sequence $a_1, a_2, \ldots$ in $\mathbb{R}$, $\{a\}$ and $\{a_1, a_2, \ldots\}$ are all Borel sets. In fact, we have

$$\{a\} = [a-1, a] \cap [a, a+1] \in \mathcal{B}, \quad \text{and} \quad \{a_1, a_2, \ldots\} = \bigcup_n \{a_n\} \in \mathcal{B}.$$

Definition 1.2

Let $\mathcal{F}$ be a σ-field on Ω. A *probability measure P* is a function

$$P : \mathcal{F} \to [0,1]$$

such that

1) $P(\Omega) = 1$; i.e. $P(\emptyset) = 0$
2) if $A_1, A_2, \ldots$ are pairwise disjoint sets (that is, $A_i \cap A_j = \emptyset$ for $i \neq j$) belonging to $\mathcal{F}$, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots .$$

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*. The [1] sets belonging to $\mathcal{F}$ are called *events*. An event A is said to occur [2] *almost surely* (a.s.) whenever $P(A) = 1$. [3] an event A is said to be a P-null event if $P(A) = 0$

Example 1.2

We take the unit interval $\Omega = [0,1]$ with the σ-field $\mathcal{F} = \mathcal{B}([0,1])$ of Borel sets $B \subset [0,1]$, and *Lebesgue measure* $P = \mathrm{Leb}$ on $[0,1]$. Then $(\Omega, \mathcal{F}, P)$ is a probability space. Recall that Leb is the unique measure defined on Borel sets such that

$$\mathrm{Leb}[a,b] = b - a$$

for any interval $[a,b]$. (In fact Leb can be extended to a larger σ-field, but we shall need Borel sets only.)

$$\Rightarrow \begin{cases} \mathrm{Leb}(\{a\}) = a - a = 0 \\ \Downarrow \\ \mathrm{Leb}(\{a_1, a_2 \ldots\}) = \sum \mathrm{Leb}(\{a_n\}) = \sum 0 = 0 \end{cases}$$

A Probability Space: $(\Omega; F; P)$

Ω: The set of all possible outcomes of the experiment.

F: A typical event / a candidate of the outcome

P: A function that measures the size of an event. / probability.

Example:

Sampling from an unbiased random number generator, it will generate an natural number between 1 and 7. Define an experiment, which uses this generator 3 times to provide a sequence of 3 consecutive (连续) numbers.

$\Rightarrow$

$\Omega = \{w = (a,b,c) \mid a,b,c \in \{1, \ldots 7\}\}$

$F = \{\text{Event} \subset \Omega\}$

$P: F \to [0,1]$ $\quad P(E) = \dfrac{\#E}{\#\Omega}$

Exercise 1.1

Show that if $A_1, A_2, \ldots$ is an *expanding* sequence of events, that is,

$$A_1 \subset A_2 \subset \cdots ,$$

then

$$P(A_1 \cup A_2 \cup \cdots) = \lim_{n\to\infty} P(A_n).$$

Similarly, if $A_1, A_2, \ldots$ is a *contracting* sequence of events, that is,

$$A_1 \supset A_2 \supset \cdots ,$$

then

$$P(A_1 \cap A_2 \cap \cdots) = \lim_{n\to\infty} P(A_n).$$

(1) Proof: If $A_1 \subset A_2 \subset \cdots$, then transfer it to pairwise disjoint version: $A_1 \cup A_2 \cup \cdots = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots$

By the definition 1.2,

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots) &= P[A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots] \\ &= P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \cdots \\ &= \lim_{n\to\infty} [P(A_1) + P(A_2 \setminus A_1) + \cdots + P(A_n \setminus A_{n-1})] \\ &= \lim_{n\to\infty} [A_1 \cup (A_2 \setminus A_1) \cup \cdots \cup (A_n \setminus A_{n-1})] \\ &= \lim_{n\to\infty} [A_1 \cup A_2 \cup \cdots \cup A_n] = \lim_{n\to\infty} P(A_n) \end{aligned}$$

the lase equality holds as $P(A_1) + P(A_2 \setminus A_1) + \cdots + P(A_n \setminus A_{n-1}) = P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_n)$

Q.E.D.

(2) Proof: If $A_1 \supset A_2 \supset \cdots$ then $A_1, A_2, \cdots$ is a contracting sequence of events, then $\Omega \setminus A_1, \Omega \setminus A_2 \cdots$ is an expanding sequence of events.

Also, as $\Omega \setminus (A_1 \cap A_2 \cap \cdots) = (\Omega \setminus A_1) \cup (\Omega \setminus A_2) \cup \cdots$

so by (1), $P(\Omega \setminus (A_1 \cap A_2 \cap \cdots)) = P[(\Omega \setminus A_1) \cup (\Omega \setminus A_2) \cup \cdots] = \lim_{n\to\infty} P(\Omega \setminus A_n)$

$\Rightarrow P(\Omega) - P(A_1 \cap A_2 \cap \cdots) = P(\Omega) - \lim_{n\to\infty} P(A_n)$

$P(A_1 \cap A_2 \cap \cdots) = \lim_{n\to\infty} P(A_n)$

Q.E.D.

Solution. Since $A_1 \subset A_2 \subset \cdots$, we have

$$A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots \cup (A_n \setminus A_{n-1}) \cup \cdots .$$

Note that $A_1, (A_2 \setminus A_1), \ldots (A_n \setminus A_{n-1}) \ldots$ are pairwise disjoint. Therefore, by the definition of probability measure, we have

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup \cdots) &= \mathbb{P}(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_3 \setminus A_2) + \cdots \\ &= \lim_{n\to\infty} \big(\mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \cdots + \mathbb{P}(A_n \setminus A_{n-1})\big) \\ &= \lim_{n\to\infty} \mathbb{P}(A_1 \cup (A_2 \setminus A_1) \cup \cdots \cup (A_n \setminus A_{n-1})) \\ &= \lim_{n\to\infty} \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \lim_{n\to\infty} \mathbb{P}(A_n) \end{aligned}$$

The lower semi-continuity of probability is proved. □

Remark. The probability $\mathbb{P}$ is also upper semi-continuous, i.e. if $B_1, B_2, \ldots$ is a contracting sequence of events, that is, $B_1 \supset B_2 \supset \cdots$, then

$$\mathbb{P}(B_1 \cap B_2 \cap \cdots) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Solution. Let $A_n = B_n^c = \Omega \setminus B_n$ for each n. Then, since $B_1 \supset B_2 \supset \cdots$, the sequence $A_1, A_2, \ldots$ satisfies that $A_1 \subset A_2 \subset \cdots$. By applying De Morgan's lwa, we have

$$\Omega \setminus (B_1 \cap B_2 \cap \cdots) = (\Omega \setminus B_1) \cup (\Omega \setminus B_2) \cup \cdots = A_1 \cup A_2 \cup \cdots .$$

By using the lower semicontinuity we get

$$\mathbb{P}\big(\Omega \setminus (B_1 \cap B_2 \cap \cdots)\big) = \mathbb{P}(A_1 \cup A_2 \cup \cdots) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \mathbb{P}(\Omega \setminus B_n).$$

Note that

$$\mathbb{P}\big(\Omega \setminus (B_1 \cap B_2 \cap \cdots)\big) = 1 - \mathbb{P}(B_1 \cap B_2 \cap \cdots)$$

and $\mathbb{P}(\Omega \setminus B_n) = 1 - \mathbb{P}(B_n)$. We conclude that

$$\mathbb{P}(B_1 \cap B_2 \cap \cdots) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

□

① 为用严格等于先转 pairwise disjoint.

② 在加法中将"加取极限"变为"取极限个加"

③ 靠 $A_1 \cup \cdots \cup A_n = A_n$

Lemma 1.1 (Borel–Cantelli)

Let $A_1, A_2, \ldots$ be a sequence of events such that $P(A_1) + P(A_2) + \cdots \boxed{< \infty}$ and let $B_n = A_n \cup A_{n+1} \cup \cdots$. Then $P(B_1 \cap B_2 \cap \cdots) = 0$.

Solution. Since $B_n = A_n \cup A_{n+1} \cup \cdots$, we have that $B_{n+1} \subset B_n$ for each n. Thus, $B_1, B_2, \ldots$ is a contracting sequence of events. By the results of Exercise 1.1, we have

$$\mathbb{P}(B_1 \cap B_2 \cap \cdots) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}(A_n \cup A_{n+1} \cup \cdots).$$

By the properties of probability, and since the the series $\sum_{k=1}^{\infty} \mathbb{P}(A_k)$ is convergent, we get

$$\mathbb{P}(B_1 \cap B_2 \cap \cdots) \le \lim_{n\to\infty} (\mathbb{P}(A_n) + \mathbb{P}(A_{n+1}) + \cdots) = 0.$$

It follows that $\mathbb{P}(B_1 \cap B_2 \cap \cdots) = 0$. □

Ex. 1.12

Proof: As $B_1, B_2, \cdots$ is a contracting sequence of events

then $$0 \le P(B_1 \cap B_2 \cap \cdots) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} P(A_n \cup A_{n+1} \cup \cdots) \le \lim_{n\to\infty} [P(A_n) + P(A_{n+1}) + \cdots] = 0$$

As $P(A_1) + P(A_2) + \cdots = \sum_{n=1}^{\infty} P(A_n) < \infty$ which implies

then $\sum_{n=1}^{\infty} P(A_n)$ is convergent

Thus, $P(B_1 \cap B_2 \cap \cdots) = 0$

Q.E.D.

1.2 Random Variables

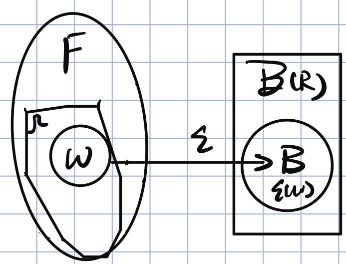
Definition 1.3

If $\mathcal{F}$ is a σ-field on Ω, then a function $\xi : \Omega \to \mathbb{R}$ is said to be *$\mathcal{F}$-measurable* if

$$\{\xi \in B\} \in \mathcal{F}$$

for every Borel set $B \in \mathcal{B}(\mathbb{R})$. If $(\Omega, \mathcal{F}, P)$ is a probability space, then such a function ξ is called a *random variable.*

F measurable function



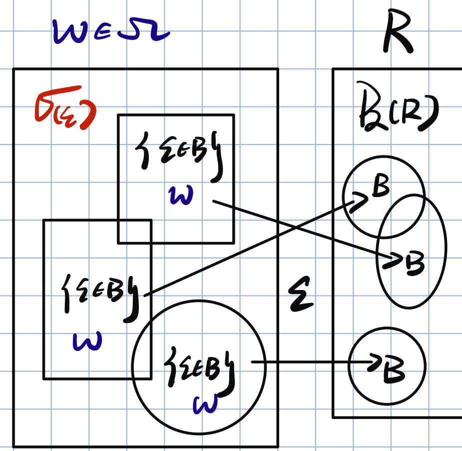
Remark 1.1

A short-hand notation for events such as $\{\xi \in B\}$ will be used to avoid clutter. To be precise, we should write

$$\{\omega \in \Omega : \xi(\omega) \in B\}$$

in place of $\{\xi \in B\}$. Incidentally, $\{\xi \in B\}$ is just a convenient way of writing the inverse image $\xi^{-1}(B)$ of a set.

- In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an $\mathcal{F}$-measurable function ξ is called a *random variable* (or simply, r.v.).



Definition 1.4

The σ-field $\sigma(\xi)$ *generated* by a random variable $\xi : \Omega \to \mathbb{R}$ consists of all sets of the form $\{\xi \in B\}$, where B is a Borel set in $\mathbb{R}$.

The σ-field $\sigma(\xi)$ represents the information we can extract about the true state ω by observing ξ.

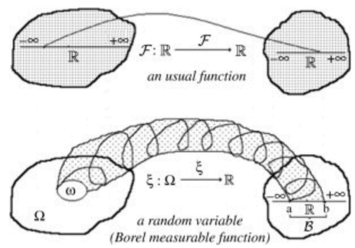
Let ξ and η be two random variables. In general, η may be not $\sigma(\xi)$-measurable since $\sigma(\xi) \subset \mathcal{F}$. If

$$\{\eta \in B\} \in \sigma(\xi)$$

for all Borel sets B in $\mathbb{R}$, η is said to be $\sigma(\xi)$-measurable. In this case, we have

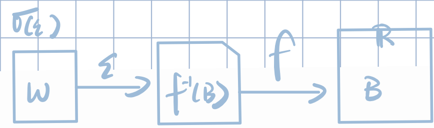
$$\sigma(\eta) \subset \sigma(\xi) \subset \mathcal{F}.$$

Remark 2.1. Fig. *2.1 illustrates the main properties of usual functions, which state correspondence between each point in $\mathbb{R}$ (argument) and some point of $\mathbb{R}$ (value function), and an F-measurable function, which state correspondence between each set B of possible values of function in $\mathbb{R}$ and some set B of corresponding realizations ω ('a random factor') from ⊗ .*



Let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be a real function. We call $\varphi(x)$ a Borel function on $\mathbb{R}$ if the inverse image $\varphi^{-1}(B)$ of any Borel set B in $\mathbb{R}$ is a Borel set.

Continuous functions are Borel functions but not all Borel functions are continuous.



Exercise 1.3

We call $f : \mathbb{R} \to \mathbb{R}$ a *Borel function* if the inverse image $f^{-1}(B)$ of any Borel set B in $\mathbb{R}$ is a Borel set. Show that if f is a Borel function and ξ is a random variable, then the composition $f(\xi)$ is $\sigma(\xi)$-measurable.

Hint Consider the event $\{f(\xi) \in B\}$, where B is an arbitrary Borel set. Can this event be written as $\{\xi \in A\}$ for some Borel set A?

Proof: As $\varphi(x)$ is a Borel function, so $\varphi^{-1}(B)$ of any Borel set B in R is a Borel set

$\Rightarrow \varphi^{-1}(B) \in \mathcal{B}(R)$

As ξ is a r.v. $\Rightarrow$ ξ is $\sigma_{(\xi)}$-measurable function,

then $\forall B \in \mathcal{B}(R)$, $\xi^{-1}(B) \in \sigma_{(\xi)}$

As $\varphi^{-1}(B)$ is a Borel set so

$\xi^{-1}(\varphi^{-1}(B)) \in \sigma_{(\xi)}$

$\Rightarrow (\xi^{-1}\varphi^{-1}) \circ (B) \in \sigma_{(\xi)}$ for $\forall B \in \mathcal{B}(R)$

$\Rightarrow \eta = \varphi(\xi)$ is $\sigma_{(\xi)}$-measurable.

Q.E.D.

Solution. Since $f(x)$ is a Borel function, and $\sigma(\xi)$ is the σ-field generated by the random variable ξ, for any $B \in \mathcal{B}(\mathbb{R})$, we have that $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$, and hence,

$$\{f(\xi) \in B\} = \{\xi \in f^{-1}(B)\} \in \sigma(\xi).$$

According to the definition of measurable functions, we conclude that the composition $f(\xi)$ is $\sigma(\xi)$-measurable. $\square$

Definition 1.5

The σ-field $\sigma\{\xi_i : i \in I\}$ generated by a family $\{\xi_i : i \in I\}$ of random variables is defined to be the smallest σ-field containing all events of the form $\{\xi_i \in B\}$, where B is a Borel set in $\mathbb{R}$ and $i \in I$.

B is any set in $\mathcal{B}(\mathbb{R})$

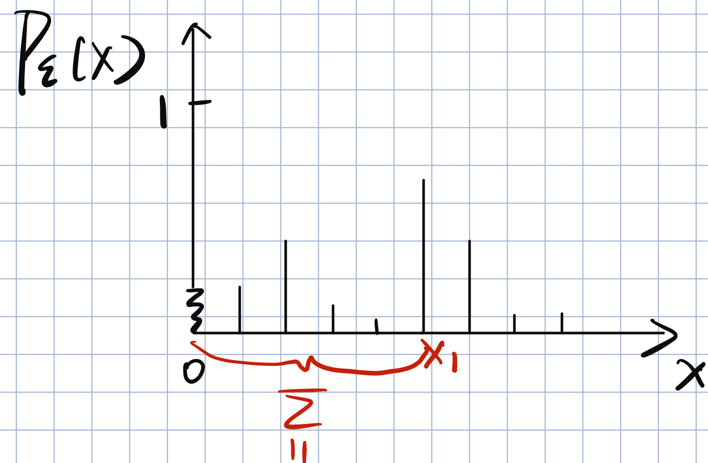
Lemma 1.2 (Doob–Dynkin)

Let ξ be a random variable. Then each $\sigma(\xi)$-measurable random variable η can be written as

$$\eta = f(\xi)$$

for some Borel function $f : \mathbb{R} \to \mathbb{R}$.

The proof of this highly non-trivial result will be omitted.



$F_\xi(x_1) = P(\{\xi \le x_1\})$.

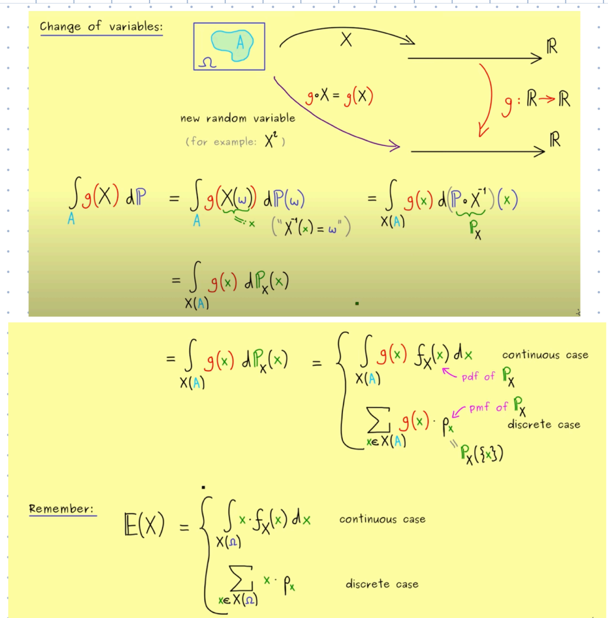
Definition 1.6

Every random variable $\xi : \Omega \to \mathbb{R}$ gives rise to a probability measure

$$P_\xi(B) = P\{\xi \in B\}$$

on $\mathbb{R}$ defined on the σ-field of Borel sets $B \in \mathcal{B}(\mathbb{R})$. We call P_ξ the *distribution* measure of ξ. The function $F_\xi : \mathbb{R} \to [0,1]$ defined by

$$F_\xi(x) = P\{\xi \le x\} \implies f_\xi(x) = \frac{d(F_\xi(x))}{dx} = P\{\xi = x\}$$



Let ξ be a r.v. then define

$$1_{\{\xi\in B\}}(\omega) = \begin{cases} 1, & \text{if } \xi(\omega) \in B \\ 0, & \text{if } \xi(\omega) \notin B \end{cases}, \quad \forall B \in \mathcal{B}(\mathbb{R})$$

Then

$$P_\xi(B) = P(\{\xi \in B\}) = \int_\Omega 1_{\{\xi\in B\}}(\omega) P(d\omega)$$

$$= \int_B 1\, P(d\omega)$$

$$\implies P_\xi(B) = \int_B 1\, dP$$

for any $B \in \mathcal{B}$, define a probability $\mathbb{P}_\xi(dx)$ on $(\mathbb{R}, \mathcal{B})$, which is called the *distribution measure* of ξ. The probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P}_\xi)$ is called the *sample space* of ξ.

Here the integral is the Lebesgue integral with respect to the probability measure $\mathbb{P}$.

The function:

$$F_\xi(x) = \mathbb{P}_\xi((-\infty, x]) = \mathbb{P}(\{\xi \le x\}), \quad x \in \mathbb{R},$$

is called the *distribution* of ξ, which possesses (Exercise 1.4):

- $F_\xi(x)$ is non-decreasing and right-continuous,
- $\lim_{x\to-\infty} F_\xi(x) = 0$ and $\lim_{x\to+\infty} F_\xi(x) = 1$.

统要

Conversely, if a function $F(x)$ having those two properties, then it must be the distribution of a random variable.

Exercise 1.4

Show that the distribution function F_ξ is non-decreasing, right-continuous, and

$$\lim_{x\to-\infty} F_\xi(x) = 0, \quad \lim_{x\to+\infty} F_\xi(x) = 1.$$

Hint For example, to verify right-continuity show that $F_\xi(x_n) \to F_\xi(x)$ for any decreasing sequence x_n such that $x_n \to x$. You may find the results of Exercise 1.1 useful.

Proof: Non-decreasing:

If $x \le y$ then $\{\xi \le x\} \subset \{\xi \le y\}$, so

$F_\xi(x) = P\{\xi \le x\} \le P\{\xi \le y\} = F_\xi(y)$

$\Rightarrow$ $F_\xi(x) \le F_\xi(y)$ if $x \le y$

$\Rightarrow$ $F_\xi(x)$ is non-decreasing.

right-continuous:

For any decreasing sequence $x_1 \ge x_2 \ge \cdots$

s.t. $\lim_{n\to\infty}(x_n) = x$ then

$\{\xi \le x_1\} \supset \{\xi \le x_2\} \supset \cdots$ form a contracting sequence

with intersection

$\{\xi \le x\} = \{\xi \le x_1\} \cap \{\xi \le x_2\} \cap \cdots = \bigcap_{n=1}^{\infty}\{\xi \le x_n\}$

then by Exercise 1.1,

$P\{\xi \le x\} = \lim_{n\to\infty}\{\xi \le x_n\}$

$\Rightarrow$ $F_\xi(x) = P\{\xi \le x\} = \lim_{n\to\infty}\{\xi \le x_n\} = \lim_{n\to\infty} F_\xi(x_n)$

Since the sequence $x_1 \ge x_2 \ge \cdots$ with $\lim_{n\to\infty} x_n = x$.

is arbitrary and $F_\xi(x)$ is non-decreasing.

Thus, $\lim_{x_n\to x^+} F_\xi(x_n) = F_\xi(x)$, F_ξ is right-continuous.

$\lim_{x\to-\infty} F_\xi(x) = 0$, $\lim_{x\to+\infty} F_\xi(x) = 1$:

Since $\{\xi \le -1\} \supset \{\xi \le -2\} \supset \cdots$ form a contracting sequence with intersection $\emptyset$; $\bigcap_{n=1}^{\infty}\{\xi \le -n\} = \emptyset$

$\{\xi \le 1\} \subset \{\xi \le 2\} \subset \cdots$ form an expanding sequence with union Ω. $\bigcup_{n=1}^{\infty}\{\xi \le n\} = \Omega$.

So by exercise 1,

Since $F_\xi(x)$ is nondecreasing

$P\{\emptyset\} = \lim_{n\to\infty} P\{\xi \le -n\} = \boxed{\lim_{n\to\infty} F_\xi(-n) = \lim_{x\to-\infty} F_\xi(x)}$

$\Rightarrow \lim_{x\to-\infty} F_\xi(x) = 0$

Since $F_\xi(x)$ is nondecreasing

$P(\Omega) = \lim_{n\to\infty} P\{\xi \le n\} = \boxed{\lim_{n\to\infty} F_\xi(n) = \lim_{x\to+\infty} F_\xi(x)}$

$\Rightarrow \lim_{x\to+\infty} F_\xi(x) = 1$

Q.E.D.

Definition 1.7

If there is a Borel function $f_\xi : \mathbb{R} \to \mathbb{R}$ such that for any Borel set $B \subset \mathbb{R}$

$$P\{\xi \in B\} = \int_B f_\xi(x)\, dx,$$

then ξ is said to be a random variable with *absolutely continuous distribution* and f_ξ is called the *density* of ξ. If there is a (finite or infinite) sequence of pairwise distinct real numbers $x_1, x_2, \ldots$ such that for any Borel set $B \subset \mathbb{R}$

$$P\{\xi \in B\} = \sum_{x_i \in B} P\{\xi = x_i\},$$

then ξ is said to have *discrete distribution* with values $x_1, x_2, \ldots$ and *mass* $P\{\xi = x_i\}$ at x_i.

Exercise 1.5

Suppose that ξ has continuous distribution with density f_ξ. Show that

$$\frac{d}{dx} F_\xi(x) = f_\xi(x)$$

if f_ξ is continuous at x.

Hint Express $F_\xi(x)$ as an integral of f_ξ.

Solution 1.5

If ξ has a density f_ξ, then the distribution function F_ξ can be written as

$$F_\xi(x) = P\{\xi \le x\} = \int_{-\infty}^{x} f_\xi(y)\, dy.$$

$\Rightarrow$ f_ξ是F_ξ的一阶导，一阶导连续 $\Rightarrow$ 原函数F_ξ可导.

Therefore, if f_ξ is continuous at x, then F_ξ is differentiable at x and

$$\frac{d}{dx} F_\xi(x) = f_\xi(x).$$

Exercise 1.6

$F_\xi(x_i) = \sum_{k=1}^{i} P(\{\xi = x_i\})$

Show that if ξ has discrete distribution with values $x_1, x_2, \ldots$, then F_ξ is constant on each interval $(s, t]$ not containing any of the x_i's and has jumps of size $P\{\xi = x_i\}$ at each x_i.

$F_\xi(x_{i+1}) - F_\xi(x_i) = P\{\xi = x_i\}$

Hint The increment $F_\xi(t) - F_\xi(s)$ is equal to the total mass of the x_i's that belong to the interval $[s, t)$.

Proof:

For $\forall s, t \in \mathbb{R}$ s.t. $s < t$, $\{\xi \le s\} \subset \{\xi \le t\}$.

$$\boxed{\{\xi \le t\} \setminus \{\xi \le s\} = \{\xi \le t\} \cap \{\xi > s\}}$$
$$= \{s < \xi \le t\} = \{\xi \in (s, t]\}.$$

If $x_i \notin (s, t]$ $\forall i \in \mathbb{N}$ then $\{\xi \in (s, t]\} = \phi$.

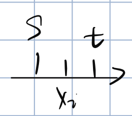
$$\Rightarrow F_\xi(t) - F_\xi(s) = P(\{\xi \le t\}) - P(\{\xi \le s\})$$
$$= P(\{\xi \le t\} \setminus \{\xi \le s\})$$
$$= P(\{s < \xi \le t\})$$
$$= P(\phi) = 0. \quad \Rightarrow F_\xi(t) = F_\xi(s)$$

As F_ξ is nondecreasing so $F_\xi(x)$ is constant on $(s, t]$.

Jump:

Assume $\exists i \in \mathbb{N}$ s.t. $x_i \in (s, t]$

As $F_\xi(t)$ is right-continuous, so



$$\lim_{t \searrow x_i} F_\xi(t) = \lim_{t \to x_i^+} F_\xi(t) = F_\xi(x_i) = P\{\xi \le x_i\}$$

$$\lim_{s \nearrow x_i} F_\xi(t) = \lim_{t \to x_i^-} F_\xi(t) = P\{\xi < x_i\}$$ * as $F_\xi(t)$ is not left-continuous.

$$\Rightarrow \lim_{t \searrow x_i} F_\xi(t) - \lim_{s \nearrow x_i} F_\xi(t) = P\{\xi \le x_i\} - P\{\xi < x_i\} = P\{\xi = x_i\}$$

Solution 1.6

If $s < t$ are real numbers such that $x_i \notin (s, t]$ for any i, then

$$F_\xi(t) - F_\xi(s) = P\{\xi \le t\} - P\{\xi \le s\} = P\{\xi \in (s, t]\} = 0,$$

i.e. $F_\xi(s) = F_\xi(t)$. Because F_ξ is non-decreasing, this means that F_ξ is constant on $(s, t]$. To show that F_ξ has a jump of size $P\{\xi = x_i\}$ at each x_i, we compute

$$\lim_{t \searrow x_i} F_\xi(t) - \lim_{s \nearrow x_i} F_\xi(s) = \lim_{t \searrow x_i} P\{\xi \le t\} - \lim_{s \nearrow x_i} P\{\xi \le s\}$$
$$= P\{\xi \le x_i\} - P\{\xi < x_i\} = P\{\xi = x_i\}.$$

* For sequence $s_1 \le s_2 \le \cdots$ with $\lim_{n \to \infty} s_n = x_i$,

$\{\xi \le s_1\} \subset \{\xi \le s_2\} \subset \cdots$ is an expanding sequence

with union $\{\xi \cancel{\le} < x_i\}$ i.e. $\{\xi < x_i\} = \bigcup_{n=1}^{\infty} \{\xi \le s_n\}$.

$$\Rightarrow P(\{\xi < x_i\}) = \lim_{n \to \infty} P(\{\xi \le s_n\}) = \lim_{n \to \infty} F_\xi(s_n) = \lim_{s_n \to x_i^-} F_\xi(s_n)$$

Definition 1.8

The *joint distribution* of several random variables $\xi_1, \ldots, \xi_n$ is a probability measure $P_{\xi_1,\ldots,\xi_n}$ on $\mathbb{R}^n$ such that

$$P_{\xi_1,\ldots,\xi_n}(B) = P\{(\xi_1, \ldots, \xi_n) \in B\}$$

for any Borel set B in $\mathbb{R}^n$. If there is a Borel function $f_{\xi_1,\ldots,\xi_n} : \mathbb{R}^n \to \mathbb{R}$ such that

$$P\{(\xi_1, \ldots, \xi_n) \in B\} = \int_B f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

for any Borel set B in $\mathbb{R}^n$, then $f_{\xi_1,\ldots,\xi_n}$ is called the *joint density* of $\xi_1, \ldots, \xi_n$.

Expectation and Moments

Definition 1.9

A random variable $\xi : \Omega \to \mathbb{R}$ is said to be *integrable* if

$$\int_\Omega |\xi|\, dP < \infty.$$ 绝对值可积

Then

$$E(\xi) = \int_\Omega \xi\, dP$$ 且积分为ξ的期望值

exists and is called the *expectation* of ξ. The family of integrable random variables $\xi : \Omega \to \mathbb{R}$ will be denoted by L^1 or, in case of possible ambiguity, by $L^1(\Omega, \mathcal{F}, P)$.

$$E[1_{A_i}] = E[1_{\{\xi \in A_i\}}]$$
$$= P(\{\xi \in A_i\})$$
$$= P_\xi(A_i)$$

Example 1.3

The *indicator function* 1_A of a set A is equal to 1 on A and 0 on the complement $\Omega \setminus A$ of A. For any event A

$$E(1_A) = \int_\Omega 1_A\, dP = P(A).$$

We say that $\eta : \Omega \to \mathbb{R}$ is a *step function* if

$$\eta = \sum_{i=1}^n \eta_i 1_{A_i},$$

where $\eta_1, \ldots, \eta_n$ are real numbers and $A_1, \ldots, A_n$ are pairwise disjoint events. Then

$$E(\eta) = \int_\Omega \eta\, dP = \sum_{i=1}^n \eta_i \int_\Omega 1_{A_i}\, dP = \sum_{i=1}^n \eta_i P(A_i).$$

$h: \mathbb{R} \to \mathbb{R}$

$E[h(\xi)]$

$= \int_{\mathbb{R}} h(x)\, dP_\xi(x)$

Exercise 1.7

Show that for any Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $h(\xi)$ is integrable

$$E(h(\xi)) = \int_{\mathbb{R}} h(x)\, dP_\xi(x).$$

Hint First verify the equality for step functions $h : \mathbb{R} \to \mathbb{R}$, then for non-negative ones by approximating them by step functions, and finally for arbitrary Borel functions by splitting them into positive and negative parts.

Solution 1.7

If h is a step function,

$$h = \sum_{i=1}^{n} h_i 1_{A_i},$$

where $h_1, \ldots, h_n$ are real numbers and $A_1, \ldots, A_n$ are pairwise disjoint Borel sets covering $\mathbb{R}$, then

$$\begin{aligned} E(h(\xi)) &= \sum_{i=1}^{n} h_i E(1_{A_i}(\xi)) = \sum_{i=1}^{n} h_i P\{\xi \in A_i\} \\ &= \sum_{i=1}^{n} h_i P_\xi(A_i) = \sum_{i=1}^{n} \int_{A_i} h(x)\, dP_\xi(x) = \int_{\mathbb{R}} h(x)\, dP_\xi(x). \end{aligned}$$

Next, any non-negative Borel function h can be approximated by a non-decreasing sequence of step functions. For such an h the result follows by the monotone convergence of integrals. Finally, this implies the desired equality for all Borel functions h, since each can be split into its positive and negative parts, $h = h^+ - h^-$, where $h^+, h^- \geq 0$.

Proof: If h is a step function, $h = \sum_{i=1}^{n} h_i I_{A_i}$

where $h_1 \cdots h_n$ are real numbers and $A_1 \cdots A_n$ are pairwise disjoint Borel sets covering R. then

$E(h(\xi)) = \sum_{i=1}^{n} h_i E(I_{A_i}(\xi)) = \sum_{i=1}^{n} h_i P\{\xi \in A_i\}$

$= \sum_{i=1}^{n} h_i P_\xi(A_i) = \sum_{i=1}^{n} \int_{A_i} h(x)\, dP_\xi(x) = \int_R h(x)\, dP_\xi(x)$

Next, as any non-negative Borel function h can be approximated by a non-decreasing sequence of step functions.

So by the Monotone Convergence of the integrals,

$E(h(\xi)) = \int_R h(x)\, dP_\xi(x)$

Finally, as all Borel functions can be split into its positive & negative parts, $h = h^+ - h^-$ where $h^+, h^- \geq 0$

then for all Borel functions $h: R \to R$,

$h(\xi)$ is integrable, $E(h(\xi)) = \int_R h(x)\, dP_\xi(x)$

Q.E.D.

假设 $h(x)$ 非负

则存在 $\{h_n(x)\}$ s.t.

h_n 非负且单增 s.t.

$h_n \to h$ when $n \to \infty$.

Assume

Secondly, from the Measure Theory or the Real Functions, we have the fact that, if $h(x)$ is a non-negative Borel function, there exists a non-decreasing sequence $\{h_n(x)\}$ of step functions such that $h_n() \to h(x)$ a.s. as $n \to \infty$. Now, assume that $h(x)$ is a non-negative Borel function such that $h(\xi)$ is integrable. Since $\{h_n(\xi)\}$ is a sequence of random variables with $0 \leq h_n(\xi) \leq h_{n+1}(\xi)$ for all n, and $h(\xi)$ is an integrable random variable such that $h_n(\xi) \to h(\xi)$ as $n \to \infty$, according to the Monotone convergence theorem and the first statement $(*)$, we have

$$\mathbb{E}[h(\xi)] = \mathbb{E}\Big[\lim_{n\to\infty} h_n(\xi)\Big] = \lim_{n\to\infty} \mathbb{E}\Big[h_n(\xi)\Big] = \lim_{n\to\infty} \int_{\mathbb{R}} h_n(x)\mathbb{P}_\xi(\mathrm{d}x).$$

Again apply the Monotone convergence theorem, we get

$$\lim_{n\to\infty} \int_{\mathbb{R}} h_n(x)\mathbb{P}_\xi(\mathrm{d}x) = \int_{\mathbb{R}} \lim_{n\to\infty} h_n(x)\mathbb{P}_\xi(\mathrm{d}x) = \int_{\mathbb{R}} h(x)\mathbb{P}_\xi(\mathrm{d}x).$$

Finally, for the general Borel function $h(x)$, we can apply the identity: $h(x) = h^+(x) - h^-(x)$, where $h^+(x)$ and $h^-(x)$ are all non-negative Borel functions defined by

$$\boxed{h^+(x) = \max\{h(x),\ 0\} \quad \text{and} \quad h^-(x) = \max\{-h(x),\ 0\},}$$

and respectively consider the cases of $h^+(x)$ and $h^-(x)$. The problem is solved. □

In particular, Exercise 1.7 implies that if ξ has an absolutely continuous distribution with density f_ξ, then

$$E(h(\xi)) = \int_{-\infty}^{+\infty} h(x) f_\xi(x)\, dx. = \int_{-\infty}^{\infty} h(x)\, d\bar{f}_\xi(x)$$

If ξ has a discrete distribution with (finitely or infinitely many) pairwise distinct values $x_1, x_2, \ldots$, then

$$E(h(\xi)) = \sum_i h(x_i) P\{\xi = x_i\}.$$

Definition 1.10

A random variable $\xi : \Omega \to \mathbb{R}$ is called *square integrable* if

$$\int_\Omega |\xi|^2\, dP < \infty.$$

Then the *variance of* ξ can be defined by

$$\operatorname{var}(\xi) = \int_\Omega (\xi - E(\xi))^2\, dP.$$

The family of square integrable random variables $\xi : \Omega \to \mathbb{R}$ will be denoted by $L^2(\Omega, \mathcal{F}, P)$ or, if no ambiguity is possible, simply by L^2.

Remark 1.2

The result in Exercise 1.8 below shows that we may write $E(\xi)$ in the definition of variance.

Exercise 1.8

Show that if ξ is a square integrable random variable, then it is integrable.

Hint Use the Schwarz inequality

$$[E(\xi\eta)]^2 \leq E(\xi^2)\, E(\eta^2) \tag{1.1}$$

with an appropriately chosen η.

L^2 is stronger than L_1 ② L^1

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) \subset L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Proof of Cauchy-Schwarz inequality. If $\mathbb{E}[\xi^2] = 0$ or $\mathbb{E}[\eta^2] = 0$, the inequality $(*)$ is trivial. We assume that $\mathbb{E}[\xi^2] > 0$ and $\mathbb{E}[\eta^2] > 0$. Using the elementary inequality that for any $a \geq 0$ and $b \geq 0$, $2ab \leq a^2 + b^2$, we have

$$\frac{|\xi\eta|}{\sqrt{\mathbb{E}[\xi^2]\mathbb{E}[\eta^2]}} = \frac{|\xi|}{\sqrt{\mathbb{E}[\xi^2]}} \cdot \frac{|\eta|}{\sqrt{\mathbb{E}[\eta^2]}} \leq \frac{|\xi|^2}{2\mathbb{E}[\xi^2]} + \frac{|\eta|^2}{2\mathbb{E}[\eta^2]}.$$

By integrating, we get

$$\begin{aligned} \frac{\mathbb{E}[|\xi\eta|]}{\sqrt{\mathbb{E}[\xi^2]\mathbb{E}[\eta^2]}} &= \frac{1}{\sqrt{\mathbb{E}[\xi^2]\mathbb{E}[\eta^2]}} \int_\Omega |\xi(\omega)\eta(\omega)|\mathbb{P}(\mathrm{d}\omega) \\ &\leq \frac{1}{2\mathbb{E}[\xi^2]} \int_\Omega |\xi(\omega)|^2\mathbb{P}(\mathrm{d}\omega) + \frac{1}{2\mathbb{E}[\eta^2]} \int_\Omega |\eta(\omega)|^2\mathbb{P}(\mathrm{d}\omega). \end{aligned}$$

This means that $E[\xi^2] = \int_\Omega |\xi(\omega)|^2 P(d\omega)$

$$\frac{\mathbb{E}[|\xi\eta|]}{\sqrt{\mathbb{E}[\xi^2]\mathbb{E}[\eta^2]}} \leq \frac{1}{2} + \frac{1}{2} = 1,$$

or

$$\Big(\mathbb{E}[|\xi\eta|]\Big)^2 \leq \mathbb{E}[\xi^2]\mathbb{E}[\eta^2].$$

Thus, the Cauchy-Schwarz inequality $(*)$ holds. □

Proof: Assume ξ is square integrable and random variable $\eta = 1$,

then $[E(|\xi|)]^2 = [E(\eta|\xi|)]^2 \leq E(\eta^2) \cdot E(|\xi|^2) = E(1) \cdot E(\xi^2) = E(\xi^2) < \infty$

the inequality is derived by the Schwarz inequality.

So $E(|\xi|) = \int_\Omega |\xi(\omega)|\, P(d\omega) < \infty$

$\Rightarrow$ ξ is integrable.

Q.E.D.

Exercise 1.9

Show that if $\eta : \Omega \to [0, \infty)$ is a non-negative square integrable random variable, then

$$E(\eta^2) = 2\int_0^\infty tP(\eta > t)\,dt.$$

Hint Express $E(\eta^2)$ in terms of the distribution function $F_\eta(t)$ of η and then integrate by parts.

Prof. Let $F_\eta(t) = P\{\eta \le t\}$ be the distribution function of η,

$\Rightarrow \quad P\{\eta > t\} = 1 - F_\eta(t)$ Let $h(\eta) = \eta^2$ then $E[\eta^2] = E[h(\eta)]$

$= \int_{\mathbb{R}} h(x)\,dP_\eta(x)$

then as η is square integrable,

$$\boxed{E(\eta^2) = \int_0^\infty t^2\,dF_\eta(t)} = \lim_{a\to+\infty}\int_0^a t^2\,dF_\eta(t)$$

$E(x^2) = \int_0^\infty x^2\,dF(x).$

$= \int_0^\infty x^2 f(x)\,dx.$

\# 为之后创造 $P\{\eta > t\}$

$$= \lim_{a\to+\infty}\int_0^a t^2\,d(F_\eta(t) - 1)$$

$$= \lim_{a\to\infty}\Big[t^2F_\eta(t) - t^2\Big]_0^a - \lim_{a\to\infty}\int_0^a (F_\eta(t) - 1)\,dt^2$$

$$= \lim_{a\to\infty} a^2(F_\eta(a) - 1) - \lim_{a\to\infty} 2\int_0^a t(F_\eta(t) - 1)\,dt$$

$$= \lim_{a\to\infty} -a^2(1 - F_\eta(a)) + \lim_{a\to\infty} 2\int_0^a t(1 - F_\eta(t))\,dt$$

As $\lim_{a\to+\infty} F_\eta(a) = 1$ so $\lim_{a\to+\infty} a^2(F_\eta(a) - 1) \to 0.$

So $E(\eta^2) = 2\int_0^\infty t\,P\{\eta > t\}\,dt$

Q.E.D.

Basic properties of expectation.

(1) If $c_1, \ldots, c_n$ are constants, then

$$\mathbb{E}\Big[\sum_{i=1}^n c_i\xi_i\Big] = \sum_{i=1}^n c_i\mathbb{E}[\xi_i].$$

(2) If $\xi \ge 0$, then $\mathbb{E}[\xi] \ge 0$; and if $\xi = 0$, then $\mathbb{E}[\xi] = 0$.

(3) If $\xi \equiv c$ for some $c \in \mathbb{R}$, then $\mathbb{E}[\xi] = c$.

(4) If $A \in \mathcal{F}$, then $\mathbb{E}[1_A] = \mathbb{P}(A)$.

(5) **Monotone convergence theorem**. If $\{\xi_n\}$ is a sequence of random variables with $0 \le \xi_n \le \xi_{n+1}$ for all n, and there is an integrable random variable ξ such that $\xi_n \to \xi$ as $n \to \infty$, then

$$\lim_{n\to\infty}\mathbb{E}[\xi_n] = \mathbb{E}\Big[\lim_{n\to\infty}\xi_n\Big] = \mathbb{E}[\xi]. \qquad (*)$$

(6) **Lebesgue's dominated convergence theorem**. If $\{\xi_n\}$ is a sequence of random variables such that $\xi_n \to \xi$ as $n \to \infty$, and there exists an integrable random variable η such that $|\xi_n| \le |\eta|$ holds for all n, then $(*)$ holds.

$\{f_n\}_1^\infty$: $0 \le f_n \le f_{n+1} \le \cdots$ $\forall n \in N$

$\exists f$ s.t. $f_n \to f$ as $n \to \infty$ then

$\lim_{n\to\infty} E[f_n] = E[\lim_{n\to\infty} f_n] = E[f]$

Let ξ be a random variable with $\mathbb{E}[\xi] = \mu$ and $\mathrm{var}(\xi) = \sigma^2$.

- The *skewness*[16] of ξ is defined:

$$s(\xi) = \mathbb{E}\Big[\frac{(\xi - \mu)^3}{\sigma^3}\Big].$$

- The (*excess*) *kurtosis*[17] of ξ is defined by

$$\kappa(\xi) = \mathbb{E}\Big[\frac{(\xi - \mu)^4}{\sigma^4}\Big].$$

Example. (The moments of standard normal distribution). Let ξ be a random variable having the standard normal distribution, then its mean $\mu = 0$, variance $\sigma^2 = 1$, skewness $s(\xi) = 0$ and kurtosis $\kappa(\xi) = 3$.

Let η be a random variable. We say that the distribution of η has a flatter top, if its kurtosis is less than 3 ($\kappa(\eta) < 3$); and if the distribution of η has a high peak, if its kurtosis is greater than 3 ($\kappa(\eta) > 3$). □

The function:

$$F_\xi(x) = \mathbb{P}_\xi((-\infty, x]) = \mathbb{P}(\{\xi \le x\}), \quad x \in \mathbb{R},$$

is called the *distribution of* ξ, which possesses (Exercise 1.4):

(7) **Fatou's lemma**[18]. Assume that $\{\xi_n\}$ is a sequence of nonnegative random variables, and there is an integrable random variable η such that $\xi_n \le \eta$ for all n. Then

$$\mathbb{E}\Big[\liminf_n \xi_n\Big] \le \liminf_n \mathbb{E}[\xi_n] \le \limsup_n \mathbb{E}[\xi_n] \le \mathbb{E}\Big[\limsup_n \xi_n\Big]$$

Remark. Let $x_1, x_2, \ldots$ be a sequence of real numbers. Then, the limit inferior and limit superior are respectively by

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty}\Big(\inf_{k\ge n} x_k\Big) \quad \text{and} \quad \limsup_{n\to\infty} x_n = \lim_{n\to\infty}\Big(\sup_{k\ge n} x_k\Big)$$

In general, the relationship of limit inferior and limit superior for sequences of real numbers is as follows:

$$\inf_n x_n \le \liminf_{n\to\infty} x_n \le \limsup_{n\to\infty} x_n \le \sup_n x_n,$$

and the limit $\lim_{n\to\infty} x_n$ exists if and only if

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = \lim_{n\to\infty} x_n.$$

Assume that $\mathbb{E}[|\xi||\eta|] < \infty$. Then, the *covariance* of ξ and η is defined by

$$\mathrm{cov}(\xi, \eta) = \mathbb{E}\Big[(\xi - \mathbb{E}[\xi])(\eta - \mathbb{E}[\eta])\Big].$$

The covariance $\mathrm{cov}(\xi, \eta)$ measures the strength of the linear relationship between ξ and η. If $\mathrm{cov}(\xi, \eta) = 0$, then ξ and η are said to be *uncorrelated*.

1.3 Conditional Probability and Independence

Definition 1.11

For any events $A, B \in \mathcal{F}$ such that $P(B) \neq 0$ the *conditional probability* of A given B is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Exercise 1.10

Prove the *total probability formula*

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots$$

for any event $A \in \mathcal{F}$ and any sequence of pairwise disjoint events $B_1, B_2, \ldots \in \mathcal{F}$ such that $B_1 \cup B_2 \cup \cdots = \Omega$ and $P(B_n) \neq 0$ for any n.

Hint $A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots$.

Since $B_1 \cup B_2 \cup \cdots = \Omega$, so $A = A \cap (B_1 \cup B_2 \cup \cdots)$

$= (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cdots$

where $(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \phi = \phi$

$\Rightarrow A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots$

$\Rightarrow P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots$

$= P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \cdots$

as $P(A|B_i) = \dfrac{P(A \cap B_i)}{P(B_i)}$ $\forall i \in \mathbb{N}$.

Q.E.D.

Solution 1.10

Since $B_1 \cup B_2 \cup \cdots = \Omega$,

$$A = A \cap (B_1 \cup B_2 \cup \cdots) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots,$$

where

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset.$$

By countable additivity

$$\begin{aligned} P(A) &= P((A \cap B_1) \cup (A \cap B_2) \cup \cdots) \\ &= P(A \cap B_1) + P(A \cap B_2) + \cdots \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots. \end{aligned}$$

Definition 1.12

Two events $A, B \in \mathcal{F}$ are called *independent* if

$$P(A \cap B) = P(A)P(B).$$

In general, we say that n events $A_1, \ldots, A_n \in \mathcal{F}$ are *independent* if

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

for any indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

Exercise 1.11

Let $P(B) \neq 0$. Show that A and B are independent events if and only if $P(A|B) = P(A)$.

Hint If $P(B) \neq 0$, then you can divide by it.

Solution 1.11

If $P(B) \neq 0$, then A and B are independent if and only if

$$P(A) = \frac{P(A \cap B)}{P(B)}.$$

In turn, this equality holds if and only if $P(A) = P(A|B)$.

Variable Independence

Definition 1.13

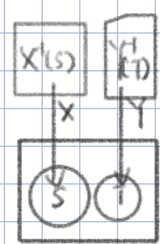
Two random variables ξ and η are called *independent* if for any Borel sets $A, B \in \mathcal{B}(\mathbb{R})$ the two events

$$\{\xi \in A\} \quad \text{and} \quad \{\eta \in B\}$$

are independent. We say that n random variables $\xi_1, \ldots, \xi_n$ are *independent* if for any Borel sets $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ the events

$$\{\xi_1 \in B_1\}, \ldots, \{\xi_n \in B_n\}$$

are independent. In general, a (finite or infinite) family of random variables is said to be *independent* if any finite number of random variables from this family are independent.



Proposition 1.1

If two integrable random variables ξ and η are independent, then

$$\mathbb{E}[\xi\eta] = \mathbb{E}[\xi]\mathbb{E}[\eta],$$

provided that the product $\xi\eta$ is also integrable. If $\xi_1, \ldots, \xi_n$ are independent integrable random variables, then

$$\mathbb{E}[\xi_1\xi_2 \cdots \xi_n] = \mathbb{E}[\xi_1]\mathbb{E}[\xi_2] \cdots \mathbb{E}[\xi_n],$$

provided that the product $\xi_1\xi_2 \cdots \xi_n$ is also integrable.

Corollary of Proposition 1.1

Assume that two integrable random variables ξ and η are independent, and the product $\xi\eta$ is integrable. Then ξ and η are uncorrelated, i.e. $\text{cov}(\xi, \eta) = 0$.

Remark. Conversely, two random variables ξ and η being uncorrelated does not imply that ξ and η must be independent.

Two random variables ξ and η:

independent $\underset{\not\leftarrow}{\rightarrow}$ uncorrelated

Proposition (Independence of two normal random variables)

Let (ξ_1, ξ_2) be a jointly normally distributed random variables. Then, ξ_1 and ξ_2 are independent if and only if they are uncorrelated, i.e. $\text{cov}(\xi_1, \xi_2) = 0$.

Proposition 1.1

If two integrable random variables $\xi, \eta : \Omega \to \mathbb{R}$ are independent, then they are *uncorrelated*, i.e.

$$E(\xi\eta) = E(\xi)E(\eta),$$

provided that the product $\xi\eta$ is also integrable. If $\xi_1, \ldots, \xi_n : \Omega \to \mathbb{R}$ are independent integrable random variables, then

$$E(\xi_1\xi_2\cdots\xi_n) = E(\xi_1)E(\xi_2)\cdots E(\xi_n),$$

provided that the product $\xi_1\xi_2\cdots\xi_n$ is also integrable.

independent $\not\Leftrightarrow$ uncorrelated

Proposition (Independence of two normal random variables)

Let (ξ_1, ξ_2) be a jointly normally distributed random variables. Then, ξ_1 and ξ_2 are independent if and only if they are uncorrelated, i.e. $\mathrm{cov}(\xi_1, \xi_2) = 0$.

Field Independence

Definition 1.14

Two σ-fields $\mathcal{G}$ and $\mathcal{H}$ contained in $\mathcal{F}$ are called *independent* if any two events

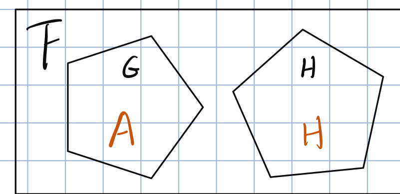
$$A \in \mathcal{G} \quad \text{and} \quad B \in \mathcal{H}$$

are independent. Similarly, any finite number of σ-fields $\mathcal{G}_1, \ldots, \mathcal{G}_n$ contained in $\mathcal{F}$ are *independent* if any n events

$$A_1 \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n$$

are independent. In general, a (finite or infinite) family of σ-fields is said to be *independent* if any finite number of them are independent.

\# Variable & σ-field 的 independence 都落到 events



Exercise 1.12

Show that two random variables ξ and η are independent if and only if the σ-fields $\sigma(\xi)$ and $\sigma(\eta)$ generated by them are independent.

Hint The events in $\sigma(\xi)$ and $\sigma(\eta)$ are of the form $\{\xi \in A\}$, and $\{\eta \in B\}$, where A and B are Borel sets.

Sometimes it is convenient to talk of independence for a combination of random variables and σ-fields.

Solution 1.12

The σ-fields $\sigma(\xi)$ and $\sigma(\eta)$ consist, respectively, of events of the form

$$\{\xi \in A\} \quad \text{and} \quad \{\eta \in B\},$$

where A and B are Borel sets in $\mathbb{R}$. Therefore, $\sigma(\xi)$ and $\sigma(\eta)$ are independent if and only if the events $\{\xi \in A\}$, and $\{\eta \in B\}$ are independent for any Borel sets A and B, which in turn is equivalent to ξ and η being independent.

Definition 1.15

We say that a random variable ξ is *independent* of a σ-field $\mathcal{G}$ if the σ-fields

$$\sigma(\xi) \quad \text{and} \quad \mathcal{G}$$

are independent. This can be extended to any (finite or infinite) family consisting of random variables or σ-fields or a combination of them both. Namely, such a family is called *independent* if for any finite number of random variables $\xi_1, \ldots, \xi_m$ and σ-fields $\mathcal{G}_1, \ldots, \mathcal{G}_n$ from this family the σ-fields

$$\sigma(\xi_1), \ldots, \sigma(\xi_m), \mathcal{G}_1, \ldots, \mathcal{G}_n$$

are independent.

Introduction to Stochastic Process (MATH3002)

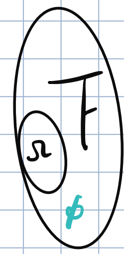
A Review of Chapters 1 and 2

1. σ-Fields

Let X be a set. In general, we use the calligraphy type of English letters $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{F}$, etc. to express the families (sets) whose elements are some subsets of X. For example, letting A, B, and C be subsets of X, $\mathcal{A} = \{A, B, C\}$ is a family.

σ-Field: Let Ω be a nonempty set. A family $\mathcal{F}$ of subsets of Ω is called a *σ-algebra* (or, *σ-field*) on Ω if it satisfies the following three properties:

- Ω belongs to $\mathcal{F}$.
- $\mathcal{F}$ is closed under the complement, i.e., if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
- $\mathcal{F}$ is closed under the countable unions, i.e., if $\{A_n\}$ is a sequence of sets in $\mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.



Using the De Morgan's law, we can show that the following statements hold:

Proposition 1

Let $\mathcal{F}$ be σ-field on Ω.

(1) $\emptyset, \Omega \in \mathcal{F}$.

(2) $\mathcal{F}$ is closed under the difference, i.e., if $A, B \in \mathcal{F}$, then $A \setminus B$ is in $\mathcal{F}$.

(3) $\mathcal{F}$ is closed under the finite unions and intersections, i.e., if $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then $\bigcap_{i=1}^{n} A_i$ and $\bigcup_{i=1}^{n} A_i$ are all in $\mathcal{F}$.

(4) $\mathcal{F}$ is closed under the countable intersections, i.e., if $\{A_n\}$ is a sequence of sets in $\mathcal{F}$, then $\bigcap_{n=1}^{\infty} A_n$ is in $\mathcal{F}$.

F may not be closed under countable (infinite) unions.

⇓

不能因为G中 Countably infinite 个B∈G都满足某性质推得 G即满足某性质

Using the De Morgan's law, we can show that the following statements hold:

Proposition 1

Let $\mathcal{F}$ be σ-field on Ω.

(1) $\emptyset, \Omega \in \mathcal{F}$.

(2) $\mathcal{F}$ is closed under the difference, i.e., if $A, B \in \mathcal{F}$, then $A \setminus B$ is in $\mathcal{F}$.

(3) $\mathcal{F}$ is closed under the finite unions and intersections, i.e., if $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then $\bigcap_{i=1}^{n} A_i$ and $\bigcup_{i=1}^{n} A_i$ are all in $\mathcal{F}$.

(4) $\mathcal{F}$ is closed under the countable intersections, i.e., if $\{A_n\}$ is a sequence of sets in $\mathcal{F}$, then $\bigcap_{n=1}^{\infty} A_n$ is in $\mathcal{F}$.

Proof. (1) According to the definition of σ-field, we have

$$\emptyset \in \mathcal{F} \Longrightarrow \Omega = \emptyset^c \in \mathcal{F}.$$

(2) If $A_1, A_2, \ldots, A_n \in \mathcal{F}$, letting $A_{n+k} = \emptyset$ for $k = 1, 2, \ldots$, then, according to the definition of σ-field, we have

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Applying the above result and according to the definition of σ-field, we have

$$A_1, \ldots, A_n \in \mathcal{F} \Longrightarrow A_1^c, \ldots, A_n^c \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{n} A_i^c \in \mathcal{F}.$$

Now, employing the de Morgan laws, we get

$$A_1, \ldots, A_n \in \mathcal{F} \Longrightarrow \bigcap_{i=1}^{n} A_i = \left(\bigcup_{i=1}^{n} A_i^c\right)^c \in \mathcal{F}.$$

(3) Using the results in (2) and according to the definition of σ-field we have

$$A, B \in \mathcal{F} \Longrightarrow A, B^c \in \mathcal{F} \Longrightarrow A \setminus B = A \cap B^c \in \mathcal{F}.$$

(4) According to the definition of σ-field, we have that, if $\{A_n\}$ is a sequence of sets in $\mathcal{F}$, then $\{A_n^c\}$ is also a sequence of sets in $\mathcal{F}$. Again applying the the de Morgan laws and according to the definition of σ-field, we obtain

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{F}.$$

The proposition is proved. $\square$

Proposition 2

(1) If $\mathcal{F}$ and $\mathcal{G}$ are two σ-fields on Ω, then $\mathcal{F} \cap \mathcal{G}$ is a σ-field on Ω.

(2) If $\mathcal{F}_i$, $i \in I$, is a family of σ-fields on Ω, then $\bigcap_{i \in I} \mathcal{F}_i$ is a σ-field on Ω.

$\Rightarrow$ 在Ω上的σ-field的交集还是在Ω上的σ-field.

2. Borel σ-Field

Borel σ-field: Let $\mathbb{R}$ be the set of real numbers. Denote $\mathcal{B}(\mathbb{R})$ (or, $\mathcal{B}$) the smallest σ-field containing all intervals in $\mathbb{R}$,

$$[a, b],\ [a, b),\ (a, b),\ (a, \infty) \in \mathcal{B},$$

can be generated by all kinds of sets

for any $a, b \in \mathbb{R}$. Then, it is called the Borel σ-field on $\mathbb{R}$, and a set $A \in \mathcal{B}$ is called a *Borel set*.

Here the smallest σ-field containing all intervals in $\mathbb{R}$ means that, $\mathcal{B}(\mathbb{R})$ is a σ-field containing all intervals in $\mathbb{R}$, and if $\mathcal{A}$ is a σ-field containing all intervals in $\mathbb{R}$ then $\mathcal{B}(\mathbb{R}) \subset \mathcal{A}$, or, letting $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ ... are all σ-fields containing all intervals in $\mathbb{R}$, then

$$\mathcal{B}(\mathbb{R}) = \mathcal{A} \cap \mathcal{B} \cap \mathcal{C} \cap \cdots .$$

From the above discussion, we can verify the following statements:

(1) All intervals in $\mathbb{R}$, such as $[a, b]$, $[a, b)$, (a, b), (a, ∞), are Borel sets;

(2) For any real number $a \in \mathbb{R}$ and any sequence $a_1, a_2, \ldots$ in $\mathbb{R}$, $\{a\}$ and $\{a_1, a_2, \ldots\}$ are Borel sets;

(3) Any open sets and any closed sets in $\mathbb{R}$ are Borel sets.

3 Foundation of Probability

Probability: Let $\mathcal{F}$ be a σ-field on Ω. A set function $\mathbb{P}(A)$ on $\mathcal{F}$ is called to be a *probability*) if it possesses:

- $\mathbb{P}(\Omega) = 1$, i.e. $\mathbb{P}(\emptyset) = 0$;
- $\mathbb{P}(\cdot)$ is σ-additive, i.e. if $A_1, A_2, \ldots$ are pairwise disjoint (i.e, $A_i \cap A_j = \emptyset$ for $i \neq j$) sets belonging to $\mathcal{F}$, then

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Using the definition we can prove the basic rules of probability:

(1) For any $A \in \mathcal{F}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(2) $\mathbb{P}(\emptyset) = 0$.

(3) If $A, B \in \mathcal{F}$ and $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(4) For all $A \in \mathcal{F}$, $0 \leq \mathbb{P}(A) \leq 1$.

(5) For any $A, B \in \mathcal{F}$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

(6) ($\mathbb{P}(\cdot)$ is sub-additive) If $A_1, A_2, \ldots$ are sets belonging to $\mathcal{F}$, then

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

(7) Let $A_1, A_2, \ldots$ be an *expanding* sequence of events, that is, $A_1 \subset A_2 \subset \cdots$. Then,

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(8) Let $A_1, A_2, \ldots$ be a *contracting* sequence of events, that is, $A_1 \supset A_2 \supset \cdots$. Then,

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

4. Measurability of Random Variables

Let Ω be a non-empty set. A mapping $\xi : \Omega \longrightarrow \mathbb{R}$ is called a function from ω into $\mathbb{R}$ if for every $\omega \in \Omega$ there exists a unique $y \in \mathbb{R}$ such that $\xi(\omega) = y$. Let $B \subset \mathbb{R}$. The *inverse image* of B under ξ is the set $\xi^{-1}(B)$ (or, $\{\xi \in B\}$) defined by

$$\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\}.$$

For the inverse image, we have the following relations

- $\xi^{-1}\big(\bigcup_{i\in I} B_i\big) = \bigcup_{i\in I} \xi^{-1}(B_i)$;
- $\xi^{-1}\big(\bigcap_{i\in I} B_i\big) = \bigcap_{i\in I} \xi^{-1}(B_i)$;
- $\xi^{-1}(B^c) = \big(\xi^{-1}(B)\big)^c$.

Here $\{B_i : i \in I\}$ is a family of subsets of $\mathbb{R}$, I is an index set, $B \subset \mathbb{R}$, and ξ be a a function from ω into $\mathbb{R}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $\xi : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$-measurable if for each $B \in \mathcal{B}(\mathbb{R})$, the set:

$$\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\} = \{\xi \in B\} \in \mathcal{F}.$$

An $\mathcal{F}$-measurable function ξ is called a *random variable*.

- Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$. A function ξ is said to be $\mathcal{G}$-measurable if and only if $\{\xi \in B\} \in \mathcal{G}$ for all $B \in \mathrm{B}(\mathbb{R})$. A $\mathcal{G}$-measurable function ξ must be $\mathcal{F}$-measurable, and an $\mathcal{F}$-measurable function may be not $\mathcal{G}$-measurable.
- Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$, and $A \in \mathcal{G}$. Then, 1_A is $\mathcal{G}$-measurable. $\longrightarrow$ 善用indicator function作为所证的特殊情况.
- Let $\mathcal{G} = \{\emptyset, \Omega\}$. A random variable ξ is $\mathcal{G}$-measurable if and only if ξ is constant, i.e. $\xi = c$ a.e.

$\xi = \xi^+ - \xi^-$

$\xi = 1_A \implies \underbrace{\xi = c_i 1_A}_{\text{step}} \implies \xi$ be a general function

Let $\xi : \Omega \longrightarrow \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote $\sigma(\xi)$ the σ-field generated by ξ, which is defined to be the smallest σ-field that consists of all sets of the form $\xi^{-1}(B)$, where B is a Borel set in $\mathbb{R}$.

From the definitions, we have the following facts:

- ξ is $\mathcal{F}$-measurable;
- ξ is $\sigma(\xi)$-measurable;
- $\sigma(\xi) \subset \mathcal{F}$.

A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is said to be a Borel function if the inverse image $f^{-1}(B)$ of any Borel set B in $\mathbb{R}$ is a Borel set.

we have the following two results:

- Let $\xi : \Omega \longrightarrow \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a Borel function. Define $\eta = f(\xi)$. Then, η is $\sigma(\xi)$-measurable. (Exercise 1.3)
- Let ξ and η are two random variables. If η is $\sigma(\xi)$-measurable, then there exists a Borel function $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that $\eta = f(\xi)$. (Doob-Dynkin Lemma)

Let ξ and η are two random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, we have the following facts:

- If ξ is $\sigma(\eta)$-measurable, then $\sigma(\xi) \subset \sigma(\eta)$, and there exist a Borel function φ such that $\xi = \varphi(\eta)$.
- If ξ and η are independent, then $\sigma(\xi)$ and $\sigma(\eta)$ are independent.
- In general, we have that $\sigma(\xi) \not\subset \sigma(\eta)$ and $\sigma(\eta) \not\subset \sigma(\xi)$.
- In general, the random variable $\zeta = \mathbb{E}[\xi \mid \eta]$ is $\sigma(\eta)$-measurable, and such that for any Borel function f,

$$\mathbb{E}\left[(\xi - \zeta)^2\right] \leq \mathbb{E}\left[(\xi - f(\eta))^2\right].$$

  This means that $\zeta = \mathbb{E}[\xi \mid \eta]$ is the best estimate that one can make of ξ from knowledge of η when η is known. we also say that ζ is the regression of ξ on η.

$E[\xi \mid \eta]$

The Best estimate that one can make of ξ from the knowledge of the known η

Independence:

Event $A \perp\!\!\!\perp$ Event B

$\Updownarrow$

$P(AB) = P(A) \cdot P(B)$

random variables

$\xi \perp\!\!\!\perp \eta$

if any events from $\{\xi \in A\}$ and $\{\xi \in B\}$ respectively, where A, B are any Borel set from $\mathcal{B}(\mathbb{R})$ are independent.

$\sigma(\xi) \perp\!\!\!\perp \sigma(\eta)$

$\xi \perp\!\!\!\perp \eta$.

σ-fields $\mathcal{G} \perp\!\!\!\perp \mathcal{H}$.

if

$\forall$ event $A \in \mathcal{G}$, $B \in \mathcal{H}$,

$A \perp\!\!\!\perp B$

$$E[\xi\eta] = E[\xi]E[\eta].$$

r.v. ξ B $\perp\!\!\!\perp$ σ-field $\mathcal{G}$

if

$\sigma(\xi) \perp\!\!\!\perp \mathcal{G}$.