

Chapter 8. Random Variables

Introduction

- In Chapter 5, we considered *r. v. s* defined on a countable probability space (Ω, \mathcal{A}, P)
- We now wish to consider an arbitrary abstract space, countable or not.
- If X maps Ω into a state space (F, \mathcal{F}) , then what we will often want to compute is the probability that X takes its values in a given subset of the state space.

$$X : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F})$$

- We take these subsets to be elements of the σ -algebra \mathcal{F} of subsets of F
我们将这些集合作为 σ 代数 \mathcal{F} 的子集
- Thus, we will want to compute

$$P(\omega : X(\omega) \in A) = P(X \in A) = P(X^{-1}(A))$$

Which are three equivalent ways to write the same quantity.

三种方式是等价的

- The third is enlightening: In order to write the same quantity. We need $X^{-1}(A)$ to be an element of \mathcal{A} , the σ -algebra on Ω on which P is defined.

Definition 8.1

1. Let (E, \mathcal{E}) and (F, \mathcal{F}) be two measurable space.
A function $X : E \rightarrow F$ is called **measurable** (relative to \mathcal{E} and \mathcal{F})
If $X^{-1}(\Lambda) \in \mathcal{E}$, for all $\Lambda \in \mathcal{F}$
also writes $X^{-1}(\mathcal{F}) \subset \mathcal{E}$
一个函数 X : 将 σ 代数 \mathcal{E} 中的集合 E 映射到 σ 代数 \mathcal{F} 中的集合 F
若满足: 其逆映射 X^{-1} 将所有 σ 代数 \mathcal{F} 上的元素都映射到 σ 代数 \mathcal{E} 中
则称函数 X 是可测的
2. When $(E, \mathcal{E}) = (F, \mathcal{F})$, a measurable function X is called a **random variable** (*r. v.* in short)
在定义(1)下, 如果 $(E, \mathcal{E}) = (\Omega, \mathcal{A})$, 则可测函数 X 就称为随机变量
3. When $F = \mathbb{R}$, we usually take \mathcal{F} to be the Borel σ -algebra \mathcal{B} of \mathbb{R} .
We will do this henceforth without special mention.

Theorem 8.1

Let \mathcal{C} be a class of subsets of F such that $\sigma(\mathcal{C}) = \mathcal{F}$.

In order for a function $X : E \rightarrow F$ to be measurable (w.r.t. the σ -algebra \mathcal{E} and \mathcal{F})

$$\Leftrightarrow X^{-1}(\mathcal{C}) \subset \mathcal{E}$$

令 \mathcal{C} 是集合 F 构成的类, 且 $\sigma(\mathcal{C}) = \mathcal{F}$

则 $X : E \rightarrow F$ 可测 $\Leftrightarrow X^{-1}(\mathcal{C}) \subset \mathcal{E}$

Proof.

Necessary is clear : X 可测, 由定义, 其逆映射 X^{-1} 将所有 σ 代数 \mathcal{F} 上的元素都映射到 σ 代数 \mathcal{E} 中, 则 $X^{-1}(\mathcal{C}) \subset \mathcal{E}$

Sufficiency: 由 $\Leftrightarrow X^{-1}(\mathcal{C}) \subset \mathcal{E}$ 推可测

Suppose that $X^{-1}(C) \in \mathcal{E}$ for all $C \in \mathcal{C}$, We need to show

$$X^{-1}(\Lambda) \in \mathcal{E} \text{ for all } \Lambda \in \mathcal{F}$$

First note that

- $X^{-1}(\cup_n \Lambda_n) = \cup_n X^{-1}(\Lambda_n)$

假设 $w \in X^{-1}(\cup_n \Lambda_n)$, 希望证明 $w \in \cup_n X^{-1}(\Lambda_n)$	假设 $w \in \cup_n X^{-1}(\Lambda_n)$, 希望证明 $w \in X^{-1}(\cup_n \Lambda_n)$
则 $X(w) \in \cup_n \Lambda_n$ $\exists n_0$ s.t. $X(w) \in \Lambda_{n_0}$ $w \in X^{-1}(\Lambda_{n_0}) \subset \cup_n X^{-1}(\Lambda_n)$ $\Rightarrow X^{-1}(\cup_n \Lambda_n) \subset \cup_n X^{-1}(\Lambda_n)$	$\exists n_0$ s.t. $w \in X^{-1}(\Lambda_{n_0})$ $\Rightarrow X(w) \in \Lambda_{n_0} \subset \cup_n \Lambda_n$ $\Rightarrow w \in X^{-1}(\cup_n \Lambda_n)$ $\Rightarrow \cup_n X^{-1}(\Lambda_n) \subset X^{-1}(\cup_n \Lambda_n)$

$$\Downarrow$$

$$X^{-1}(\cup_n \Lambda_n) = \cup_n X^{-1}(\Lambda_n)$$

- $X^{-1}(\cap_n \Lambda_n) = \cap_n X^{-1}(\Lambda_n)$

假设 $w \in X^{-1}(\cap_n \Lambda_n)$, 希望证明 $w \in \cap_n X^{-1}(\Lambda_n)$	假设 $w \in \cap_n X^{-1}(\Lambda_n)$, 希望证明 $w \in X^{-1}(\cap_n \Lambda_n)$
则 $X(w) \in \cap_n \Lambda_n$ $\forall n_0$ $X(w) \in \Lambda_{n_0}$ $w \in X^{-1}(\Lambda_{n_0}) \subset \cap_n X^{-1}(\Lambda_n)$ $\Rightarrow X^{-1}(\cap_n \Lambda_n) \subset \cap_n X^{-1}(\Lambda_n)$	$\forall n_0$ $w \in X^{-1}(\Lambda_{n_0})$ $\Rightarrow X(w) \in \Lambda_{n_0} \subset \cap_n \Lambda_n$ $\Rightarrow w \in X^{-1}(\cap_n \Lambda_n)$ $\Rightarrow \cap_n X^{-1}(\Lambda_n) \subset X^{-1}(\cap_n \Lambda_n)$

$$\Downarrow$$

$$X^{-1}(\cap_n \Lambda_n) = \cap_n X^{-1}(\Lambda_n)$$

- $X^{-1}(\Lambda^c) = (X^{-1}(\Lambda))^c$

假设 $w \in X^{-1}(A^c)$, 希望证明 $w \in (X^{-1}(\Lambda))^c$	假设 $w \in (X^{-1}(\Lambda))^c$, 希望证明 $w \in X^{-1}(A^c)$
$\forall w \in X^{-1}(\Lambda^c)$ $X(w) \in \Lambda^c$ $X(w) \notin \Lambda$ $w \notin X^{-1}(\Lambda)$ $\Rightarrow w \in (X^{-1}(\Lambda))^c$ $\Rightarrow X^{-1}(A^c) \subset (X^{-1}(\Lambda))^c$	$\forall w \in (X^{-1}(\Lambda))^c$ $w \notin X^{-1}(\Lambda)$ $X(w) \notin \Lambda$ $X(w) \in \Lambda^c$ $\Rightarrow w \in X^{-1}(\Lambda^c)$ $\Rightarrow (X^{-1}(\Lambda))^c \subset X^{-1}(A^c)$

$$\Downarrow$$

$$\Rightarrow X^{-1}(A^c) = (X^{-1}(\Lambda))^c$$

定义 $\mathcal{B} = \{A \in \mathcal{F} : X^{-1}(A) \in \mathcal{E}\}$ 若 $\mathcal{F} = \mathcal{B}$, 则证明完毕

则 $\mathcal{C} \subset \mathcal{B}$

由上面的证明可知, X^{-1} 满足对可列交、可列并、补集封闭

即: 若 $A_n \in \mathcal{B} \rightarrow X^{-1}(A_n) \in \mathcal{E}$ and $A_n \in \mathcal{F}$ ($\rightarrow \mathcal{B}$ 的定义)

$$\begin{cases} \bigcup_n A_n \in \mathcal{F} & \text{and } X^{-1}(\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} X^{-1}(A_n) \in \mathcal{E} & \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{B} \\ \bigcap_n A_n \in \mathcal{F} & \text{and } X^{-1}(\bigcap_{n=1}^{\infty} A_n) = \bigcap_{n=1}^{\infty} X^{-1}(A_n) \in \mathcal{E} & \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{B} \\ A_n^c \in \mathcal{F} & \text{and } X^{-1}(A_n^c) = (X^{-1}(A_n))^c \in \mathcal{E} & \Rightarrow A_n^c \in \mathcal{B} \end{cases}$$

则 \mathcal{B} 是一个 σ 代数

由于 $\sigma(\mathcal{C})$ 的定义, $\sigma(\mathcal{C})$ 是包含 \mathcal{F} 的最小 σ 代数, 而 \mathcal{B} 是包含 \mathcal{F} 的代数, 所以 $\sigma(\mathcal{C}) \subset \mathcal{B}$ (这里集合 A 和集合 F 是一样的, 都是 \mathcal{F} 里面的元素)

由于 \mathcal{B} 的定义, \mathcal{B} 里面的集合都在 \mathcal{F} 里面, 所以 $\mathcal{B} \subset \mathcal{F}$

由于题目条件, $\mathcal{F} = \sigma(\mathcal{C})$

$$\text{则 } \begin{cases} \sigma(\mathcal{C}) \subset \mathcal{B} \subset \mathcal{F} \\ \mathcal{F} = \sigma(\mathcal{C}) \end{cases} \Rightarrow \mathcal{F} = \mathcal{B}$$

$$\text{即 } \forall A \in \mathcal{F} = \mathcal{B} \Rightarrow X^{-1}(A) \in \mathcal{E}$$

Remark

We have seen that a probability measure P on \mathbb{R} is characterized by the quantities $P((-\infty, a])$. Thus the distribution measure P^X on \mathbb{R} of a $r.v.$ X should be characterized by

$$P^X((-\infty, a]) = P(X \leq a)$$

and what is perhaps surprisingly nice is that being a $r.v.$ is further characterized only by events of the form

$$\{w : X(w) \leq a\} = \{X \leq a\}$$

Indeed, what this amounts to is that a function is measurable ---- and hence a $r.v.$ ---- if and only if its distribution function is defined.

Corollary 8.1

Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$

Let (E, \mathcal{E}) be an arbitrary measurable space.

Let X, X_n be real-valued functions on E .

1. X is measurable

$$\Leftrightarrow \{X \leq a\} = \{w : X(w) \leq a\} = X^{-1}((-\infty, a]) \in \mathcal{E}$$

$$\mathcal{C} = \{(-\infty, a], a \in \mathbb{R}\} \xrightarrow{\text{The 2.1}} \mathcal{B} = \sigma(\mathcal{C})$$

$$X^{-1}(\mathcal{C}) \in \mathcal{E}$$

则由定理8.1, X 可测

$$\Leftrightarrow \{X < a\} = \{w : X(w) < a\} = X^{-1}((-\infty, a)) \in \mathcal{E}$$

令

$$\mathcal{C} = \{(-\infty, a], a \in \mathbb{R}\}$$

$$\mathcal{Y} = \{(-\infty, a), a \in \mathbb{R}\}$$

已知 $\mathcal{B} = \sigma(\mathcal{C})$

$$(-\infty, a] = \bigcup_{n=1}^{\infty} (-\infty, a + \frac{1}{n}) \in \sigma(\mathcal{Y}) \quad \text{即} \quad \mathcal{C} \subset \sigma(\mathcal{Y}) \quad \Rightarrow \sigma(\mathcal{C}) \subset \sigma(\mathcal{Y})$$

$$(-\infty, a) = \bigcup_{n=1}^{\infty} (-\infty, a - \frac{1}{n}] \in \sigma(\mathcal{C}) \quad \text{即} \quad \mathcal{Y} \subset \sigma(\mathcal{C}) \quad \Rightarrow \sigma(\mathcal{Y}) \subset \sigma(\mathcal{C})$$

$$\therefore \sigma(\mathcal{Y}) = \sigma(\mathcal{C}) = \mathcal{B}$$

$$\forall Y \in \mathcal{Y}, X^{-1}(Y) \in \mathcal{E}$$

则由定理8.1, X 可测

each $a \in \mathbb{R}$

2. If X_n are measurable

$\Rightarrow \sup X_n$ are measurable.

$\Rightarrow \inf X_n$ are measurable.

$\Rightarrow \lim_{n \rightarrow \infty} \sup X_n$ are measurable.

$\Rightarrow \lim_{n \rightarrow \infty} \inf X_n$ are measurable.

3. If X_n are measurable

If X_n converges pointwise to X

$\Rightarrow X$ is measurable

Theorem 8.2

Let X be measurable from (E, \mathcal{E}) into (F, \mathcal{F})

Let Y be measurable from (F, \mathcal{F}) into (G, \mathcal{G})

$\Rightarrow Y \circ X$ is measurable from (E, \mathcal{E}) into (G, \mathcal{G})

$\Rightarrow Y \circ X$ 表示复合函数, 内层是 X , 外层函数是 Y

$$\Rightarrow (Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A))$$

Topology

- A **topological space** is an abstract space with a collection of open sets;
拓扑空间是抽象空间，这个抽象空间是开集合的集合

A "collection of open sets" is a collection of sets such that any union of sets in the collection is also in the collection, and any finite intersection of open sets in the collection is also in the collection.

开集合的集合是集合的集合

任意在这个集合中的集合之并仍然在这个集合里面

任意在这个集合中的有限的开集合的交集仍然在这个集合里面

***** 尝试说明

无限个开集的交集是可以构成闭集的，所以在拓扑空间中只能是有限个开集的交仍然在这个集合里面。比如说在 \mathbb{R}^n 上面：

$$\bigcap_{n=1}^{\infty} (-\infty, a + \frac{1}{n}) = (-\infty, a]$$

- The collection of open sets is called the **topology** of the space.

满足上述条件的开集的集合叫做拓扑空间

一个集合上是不能讨论它的开集是什么的，但是定义了拓扑的集合就可以知道它的哪些子集是开集。

有了开集的定义之后就可以定义闭集，闭集的定义就是开集的补集。

一个集合 X ，他的全部子集的集合叫做 X 的幂集，记作 2^X

现在定义的这个集合 X 的拓扑 τ 就是这个幂集 2^X 的一个子集，并且满足下面三个性质：

1. X 和 \emptyset 都属于集合 τ
2. τ 中有限个集合的交集属于 τ
3. τ 中任意多个集合的并集属于 τ

拓扑如此定义是为了引入“开集”（拓扑的元素即开集，拓扑亦可称“开集系”）从而“连续”这样的概念。开集的关键特征是任意个（无论有限个、可数个或不可数个）开集的并仍是开集，任意有限个开集的交仍是开集。由拓扑导出的 σ 域，也称Borel域。

- An abstract definition of a **continuous function** is as follows:
Given two topological spaces (E, \mathcal{U}) and (F, \mathcal{V})
 $\rightarrow \mathcal{U}$ are the open sets of E
 $\rightarrow \mathcal{V}$ are the open sets of V
 Then a **continuous function** $f : E \rightarrow F$ is a function such that
 $f^{-1}(A) \in \mathcal{U}$ for each $A \in \mathcal{V}$
 $f^{-1}(\mathcal{V}) \subset \mathcal{U}$
 一个抽象的定义，关于连续函数，如下：
 给定两个拓扑空间 (E, \mathcal{U}) 和 (F, \mathcal{V})
 $\rightarrow \mathcal{U}$ 是 E 的开集构成的集合

→ \mathcal{V} 是 V 的开集构成的集合

则一个连续函数 $f: E \rightarrow F$ 满足下列条件:

$f^{-1}(A) \in \mathcal{U}$ 对任意的 $A \in \mathcal{V}$

$f^{-1}(\mathcal{V}) \subset \mathcal{U}$

***** 尝试说明

- The **Borel σ -algebra** of a topological space (E, \mathcal{U}) is $\mathcal{B} = \sigma(\mathcal{U})$
(The open sets do not form a σ -algebra by themselves: they are not closed under complements or under countable intersections.)
在拓扑空间 (E, \mathcal{U}) 中的博雷尔 σ 代数是 $\mathcal{B} = \sigma(\mathcal{U})$
博雷尔 σ 代数是由拓扑空间 \mathcal{U} 生成的最小 σ 代数, 即 $\sigma(\mathcal{U})$

***** 尝试说明为什么对补集和可数交不满足

补集: 开集的补集是闭集 (开集与闭集的对偶性)

可数交: 无限个开集的交集是可以构成并集的, 所以在拓扑空间中只能是有限个开集
的交仍然在这个集合里面。

$$\bigcap_{n=1}^{\infty} (-\infty, a + \frac{1}{n}) = (-\infty, a]$$

Theorem 8.3

Let (E, \mathcal{U}) and (F, \mathcal{V}) are two topological spaces

Let \mathcal{E}, \mathcal{F} be their Borel σ -algebra.

Every continuous function X from E into F is then measurable (also called "Borel").

proof.

Since $\mathcal{F} = \sigma(\mathcal{V})$

By Theorem 8.1 it suffices to show that $X^{-1}(\mathcal{V}) \subset \mathcal{E}$

But for $O \in \mathcal{V}$, we know $X^{-1}(O)$ is open and therefore in \mathcal{E} , as \mathcal{E} being the Borel σ -algebra, it contains the class \mathcal{U} of open sets of E .

$\mathcal{E} = \sigma(\mathcal{U})$

$\mathcal{F} = \sigma(\mathcal{V})$

由拓扑空间的定义, \mathcal{V} 是一个子集的集合, $\mathcal{F} = \sigma(\mathcal{V})$, 且 连续函数 $X: E \rightarrow F$

则由拓扑空间中连续函数的性质, 可以得到 $X^{-1}(O) \in \mathcal{E}$ 对任意的 $O \in \mathcal{E}$

则由拓扑空间中连续函数的性质, 可以得到 $X^{-1}(\mathcal{V}) \in \mathcal{E}$

再由定理8.1可以得知连续函数 X 是可测的 (充要条件)

Theorem 8.1

Let \mathcal{C} be a class of subsets of F such that $\sigma(\mathcal{C}) = \mathcal{F}$.

In order for a function $X: E \rightarrow F$ to be measurable (w.r.t. the σ -algebra \mathcal{E} and \mathcal{F})

$$\Leftrightarrow X^{-1}(\mathcal{C}) \subset \mathcal{E}$$

令 \mathcal{C} 是集合 F 构成的类, 且 $\sigma(\mathcal{C}) = \mathcal{F}$

则 $X: E \rightarrow F$ 可测 $\Leftrightarrow X^{-1}(\mathcal{C}) \subset \mathcal{E}$

一个抽象的定义, 关于连续函数, 如下:

给定两个拓扑空间 (E, \mathcal{U}) 和 (F, \mathcal{V})

$\rightarrow \mathcal{U}$ 是 E 的开集构成的集合

$\rightarrow \mathcal{V}$ 是 F 的开集构成的集合

则一个连续函数 $f: E \rightarrow F$ 满足下列条件:

$$f^{-1}(A) \in \mathcal{U} \text{ 对任意的 } A \in \mathcal{V}$$

$$f^{-1}(\mathcal{V}) \subset \mathcal{U}$$

Indicator function

- Recall that for a subset A of E , the **indicator function** $I_A(x)$ is defined to be

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

- The function $I_A(x)$, usually written I_A with the argument x being implicit, "indicates" whether or not a given x is in A .

函数 $I_A(x)$ 通常写作 I_A (里面的参数 x 是隐式的), "指示" 给定的 x 是否在 A 中。

Theorem 8.4

Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$

Let (E, \mathcal{E}) be any measurable space.

1. An indicator I_A on E is measurable

$$\Leftrightarrow A \in \mathcal{E}.$$

If $B \subset \mathbb{R}$, we have

$$(I_A)^{-1}(B) = \begin{cases} \emptyset & \text{if } 0 \notin B, 1 \notin B \\ A & \text{if } 0 \notin B, 1 \in B \\ A^c & \text{if } 0 \in B, 1 \notin B \\ E & \text{if } 0 \in B, 1 \in B \end{cases}$$

The result follows.

函数 I_A 可测:

I_A 将 σ 代数 \mathcal{E} 中的集合 E 映射到 σ 代数 \mathcal{F} 中的集合 F : $I_A : E \rightarrow F$

I_A 将 σ 代数 \mathcal{F} 中的集合 F 映射到 σ 代数 \mathcal{E} 中的集合 E : If $X^{-1}(\Lambda) \in \mathcal{E}$, for all $\Lambda \in \mathcal{F}$

现在将 (F, \mathcal{F}) 换为 $(\mathbb{R}, \mathcal{B})$

I_A 将 σ 代数 \mathcal{E} 中的集合 E 映射到 σ 代数 \mathcal{B} 中的集合 \mathbb{R} : $I_A : E \rightarrow \mathbb{R}$

I_A^{-1} 将 σ 代数 \mathcal{B} 中的集合 \mathbb{R} 映射到 σ 代数 \mathcal{E} 中的集合 E : If $X^{-1}(\Lambda) \in \mathcal{E}$, for all $\Lambda \in \mathcal{B}$

如果 $B \subset \mathbb{R}$

即 $I_A(E) = B$

集合 E 的形式是 $\{x : x \in A\}$ 或者 $\{x : x \notin A\}$

集合 B 里面的元素要么是 0, 要么是 1

当集合 E 是 $\{x : x \in A\}$ 这种形式的时候, $I_A(E)$ 把集合 E 映射为 $\{1\}$

当集合 E 是 $\{x : x \notin A\}$ 这种形式的时候, $I_A(E)$ 把集合 E 映射为 $\{0\}$

则 $I_A^{-1}(B) = E$

当集合 B 中不含有 0, 也不含有 1 的时候, $I_A^{-1}(B)$ 把集合 E 映射为空集, 空集在 \mathcal{E} 里面

当集合 B 中不含有 0, 含有 1 的时候, $I_A^{-1}(B)$ 把集合 E 映射为 $\{x : x \in A\}$, 即集合 A , 集合 A 在 \mathcal{E} 里面

当集合 B 中含有 0, 不含有 1 的时候, $I_A^{-1}(B)$ 把集合 E 映射为 $\{x : x \notin A\}$, 即集合 A^c , 集合 A^c 在 \mathcal{E} 里面

当集合 B 中含有 0, 含有 1 的时候, $I_A^{-1}(B)$ 把集合 E 映射为集合 E , 集合 E 在 \mathcal{E} 里面

结合函数可测的定义: I_A^{-1} 将 σ 代数 \mathcal{B} 中的集合 \mathbb{R} 映射到 σ 代数 \mathcal{E} 中的集合 E : If $X^{-1}(\Lambda) \in \mathcal{E}$, for all $\Lambda \in \mathcal{B}$

可知结论成立。

2. If X_1, \dots, X_n are real-valued measurable functions on (E, \mathcal{E}) .

If f is Borel on \mathbb{R}^n

$\Rightarrow f(X_1, \dots, X_n)$ is measurable

X_1, \dots, X_n 是在 (E, \mathcal{E}) 上的实值可测函数。

f 是 \mathbb{R}^n 上的博雷尔

$\Rightarrow f(X_1, \dots, X_n)$ 是可测的

The Borel σ -algebra \mathcal{B}^n on \mathbb{R}^n is generated by the quadrants $\prod_{i \leq n} (-\infty, a_i]$ quadrants: 象限

Theorem 2.1

The Borel σ -algebra of \mathbb{R} is generated by intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$

By the exact same proof as was used in Theorem 2.1

Let X denote the vector-valued function $X = (X_1, \dots, X_n)$.

Thus $X: E \rightarrow \mathbb{R}^n$

Then

$$X^{-1}\left(\prod_{i \leq n} (-\infty, a_i]\right) = \bigcap_{i \leq n} \{X_i \leq a_i\} \in \mathcal{E}$$

and therefore X is a measurable function from (E, \mathcal{E}) into $(\mathbb{R}^n, \mathcal{B}^n)$.

The statement (2) then follows from Theorem 8.2

3. If X, Y are measurable

$\Rightarrow X + Y$ are measurable.

$\Rightarrow XY$ are measurable.

$\Rightarrow \max(X, Y)$ are measurable.

$\Rightarrow \min(X, Y)$ are measurable.

$\Rightarrow X/Y$ are measurable.

X, Y 是可测的 $\Rightarrow f$ 是博雷尔(定理8.3) \Rightarrow 由定理8.4第二条可知 $f(X, Y)$ 是可测的
所以现在的关键问题就是证明 f 是博雷尔(定理8.3)

定理8.3中要用到连续函数的定义, 则要回到拓扑空间中第二点关于连续函数的抽象定义:

函数 f 的逆映射将其中一个拓扑空间中的所有元素都映射到另一个拓扑空间中的元素

Theorem 8.3

Let (E, \mathcal{U}) and (F, \mathcal{V}) are two topological spaces

Let \mathcal{E}, \mathcal{F} be their Borel σ -algebra.

Every continuous function X from E into F is then measurable (also called "Borel").

连续函数的定义:

An abstract definition of a continuous function is as follows:

Given two topological spaces (E, \mathcal{U}) and (F, \mathcal{V})

$\rightarrow \mathcal{U}$ are the open sets of E

$\rightarrow \mathcal{V}$ are the open sets of V

Then a continuous function $f: E \rightarrow F$ is a function such that

$f^{-1}(A) \in \mathcal{U}$ for each $A \in \mathcal{V}$

$f^{-1}(\mathcal{V}) \subset \mathcal{U}$

Note that the function $f_1: \mathbb{R}^2 \rightarrow \mathbb{R}$ given by $f_1(x, y) = x + y$ is continuous.

So also are:

$f_2(x, y) = xy$

$f_3(x, y) = \max(x, y)$

$f_4(x, y) = \min(x, y)$

The function $f_5(x, y) = \frac{x}{y}$ is continuous from $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$ into \mathbb{R} .
 Therefore (3) follows from (2) together with Theorem 8.3
 (That continuous functions are Borel measurable)

Remark

If X is a *r. v.* on (Ω, \mathcal{A}, P) , with values in (E, \mathcal{E})

Then the **distribution measure** (or **law**) of X is defined: $\forall B \in \mathcal{E}$

$$P^X(B) = P(X^{-1}(B)) = (P \circ X^{-1})(B) = P(\{w : X(w) \in B\}) = P(X \in B)$$

- All four different ways of writing $P^X(B)$ on the right side above are used in mathematics, but the most common is

$$P^X(B) = P(X \in B)$$

Where the " w " is not explicitly written but rather implicitly understood to be there.

- It avoids specifying the space Ω (Which is often difficult to construct mathematically), and simply work with Probability measures on (E, \mathcal{E}) . Sometimes P^X is also called the *image* of P by X .

用这种方法来定义概率，避免了指定状态空间 Ω （因为 Ω 通常很难在数学上构造）；

而只需要对概率测度 (E, \mathcal{E}) 进行操作。

有时也称 P^X 为 P 在 X 下的像。

Since X^{-1} commutes with taking unions and intersections

Since $X^{-1}(E) = \Omega$

We have the following result.

Theorem 8.5 : The distribution X , (or the law of X), is a probability measure on (E, \mathcal{E}) .

The distribution X : $P^X(B) = P(X^{-1}(B)) = (P \circ X^{-1})(B) = P(\{w : X(w) \in B\}) = P(X \in B)$

The distribution X : $P^X(B) = P(X \in B)$

要证明 X 是定义在 (E, \mathcal{E}) 上的概率测度, 就需要证明其在 (E, \mathcal{E}) 上满足规范性和可列可加性.

首先看规范性:

$$P^X(E) = P(X^{-1}(E)) = P(\Omega) = 1$$

再来看可列可加性:

对于任意的 \mathcal{E} 上的可列集合 $(E_n)_{n \geq 1}$, 满足两两不交 $E_n \cap E_m = \emptyset$

$$\begin{array}{ll} P^X(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} P^X(E_n) & X^{-1}(\cup_{n=1}^{\infty} E_n) = \cup_{n=1}^{\infty} X^{-1}(E_n) \\ & X^{-1}(\cap_{n=1}^{\infty} E_n) = \cap_{n=1}^{\infty} X^{-1}(E_n) \\ & X^{-1}(E^c) = (X^{-1}(E))^c \\ & \downarrow \\ & E_n \cap E_m = \emptyset \\ & X^{-1}(\emptyset) = \emptyset \\ & X^{-1}(E_n \cap E_m) = X^{-1}(E_n) \cap X^{-1}(E_m) \\ & \Rightarrow X^{-1}(E_n \cap E_m) = X^{-1}(\emptyset) = \emptyset \\ & \text{两两不交} \end{array}$$

Remark 定义了分布函数和密度函数

- When X is a real-valued $r.v.$, Its distribution P^X is a probability on \mathbb{R} , which is entirely characterized by its distribution function:

$$F_X(x) = P^X((-\infty, x]) = P(X \leq x)$$

- The function F_X is called the **cumulative distribution function of the $r.v.$ X** .
- When F_X admits a density f_X , i. e.,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \text{for all } x \in \mathbb{R}$$

We also say that the function f_X is the **probability density of the $r.v.$ X** .

- Often the cumulative distribution function is referred as a "cdf", or simply a "distribution function". Analogously, the probability density function is referred as a "pdf", or simply a "density".