Probability theory

Probability theory

Chapter 1 Set Theory

1.1. Operations of sequence of sets

There are two types of countable for set: finite and countably infinite. Intuitively for me when the sets are finite, we should use $\cup_{i=1}^{n} A$, when the sets are countably infinite, we should use $\cup_{i=1}^{\infty} A$, when the sets are countable, we should use $\cup_{i=1} A$. But most of the time, $\cup_{i=1}^{\infty} A$ stands for countable, why?

The reason is when a theorem is correct for countably infinite, it is usually correct for finite situation, so some people often use the symbol of countably infinite for both countably infinite and finite which are countable. Let's look at σ additive for example

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$

The above is definition of σ additive (countable) , and the symbol is actually just countably infinite. Let's see why it can work like this. Using the formula, we can derive additive:

$$\mu(\cup_{i=2}^{\infty} A_i \cup A_1) = \sum_{i=2}^{\infty} \mu(A_i) + \mu(A_1) = \mu(\cup_{i=2}^{\infty} A_i) + \mu(A_1)$$

$$so$$

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

By the result we see ∞ bears two meanings: ∞ itself and finite, so it actually means countable. So in the future when we see ∞ as countable in textbook, just know in that circumstance, ∞ must bear two meanings. But for myself

- $\cup_{i=1}^{n} A$, when the sets are finite
- $\cup_{i=1}^{\infty} A$, when the sets are countably infinite
- $\cup_{i=1} A$, when the sets are countable

When using sup and inf independently, set can be both finite or countably infinite. When we combine lim with sup or inf, the set we are studying is countably infinite by default.

Based on that, let's define $lim\ sup$ and $lim\ inf$

Infinitely often:

$$\lim_{n\to\infty} supA_n = \overline{\lim_{n\to\infty}} A_n = \lim_{n\to\infty}(\sup_{i\geq n} A_i) = \inf_{n\geq 1}(\sup_{i\geq n} A_i) = \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i$$

Ultimately:

$$\lim_{n\to\infty} infA_n = \underline{\lim_{n\to\infty}} A_n = \lim_{n\to\infty}(\inf_{i\geq n} A_i) = \sup_{n\geq 1}(\inf_{i\geq n} A_i) = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i$$

现在集序列为

$$C_1 = \{1, a\},\ C_2 = \{0, b\},\ C_3 = \{1, b\},\ C_4 = \{0, b\},\ C_5 = \{1, b\}\ldots$$

上极限我们从 $n = 1$ 开始，得到内层的并集为 $\{0, 1, a, b\}$ ，然后 $n = 2$ ，内层的并集为 $\{0, 1, b\}$ ，之后也是一样。全部做完之后取交集，得到上极限为 $\{0, 1, b\}$ 。

下极限我们从 $n = 1$ 开始，得到内层的交集为 $\emptyset$ ，然后 $n = 2$ ，内层的交集为 $\{b\}$ ，之后也是一样。全部做完后取并集，得到下极限为 $\{b\}$ 。

上极限关注的元素特征只有一个，元素在无限个集合中存在，它并不关心元素是否在另外无限个集合中不存在。比如说0，1，它们都在无限个集合中存在，但是也在另外无穷个集合中不存在，可是上极限不管这个依旧把它们收录进来。

下极限关注的元素特征是，第一，元素在无限个集合中存在，第二，那些元素不存在的集合仅为有限个，这里就只有b满足条件，而0，1都不满足第二个条件。

所以我们可以看到上极限肯定包含了下极限

如果两者相等，也就是 $\overline{\lim_{n\to\infty}} A_n = \underline{\lim_{n\to\infty}} A$ ，则称这个集序列极限存在，记作 $\lim_{n\to\infty} A_n$

We have to address $lim\ sup$ and $lim\ inf$ from another angle which is function

$$\lim_{n\to\infty} sup\ f_n(x) \quad \lim_{n\to\infty} inf\ f_n(x)$$

Again, the function sequence we use here is countably infinite. First, let's consider the former at x_0

$$\lim_{n\to\infty} sup\ f_n(x_0) = \lim_{n\to\infty} (\sup_{i\geq n} f_i(x_0))$$

The above can be interpreted as following:

$$\begin{gathered} max\{f_1(x_0), f_2(x_0), \dots\} \Rightarrow v_1 \\ max\{f_2(x_0), f_3(x_0), \dots\} \Rightarrow v_2 \\ max\{f_3(x_0), f_4(x_0), \dots\} \Rightarrow v_3 \\ \dots\dots \\ max\{f_n(x_0), f_{n+1}(x_0), \dots\} \Rightarrow v_n \end{gathered}$$

$$\lim_{n\to\infty} v_n$$

Let's assume the sequence $\{v_1, v_2 \dots\}$ biggest value is at $f_k(x_0)$, so
$v_1 = v_2 = \dots = v_k = f_k(x_0)$, after step k ,since the biggest value has been stepped over, maximum value of the rest must be smaller than $f_k(x_0)$ which we denote as $f_l(x_0)$. So the sequence becomes

$$\underbrace{f_k(x_0), f_k(x_0), \dots, f_k(x_0)}_{k}, \underbrace{f_l(x_0), f_l(x_0), \dots, f_l(x_0)}_{l-k}, \dots$$

it will decrease as n increases, so

$$\lim_{n\to\infty} v_n = \inf_{n\geq 1} v_n$$

hence

$$\lim_{n\to\infty} sup\ f_n(x_0) = \lim_{n\to\infty} (\sup_{i\geq n} f_i(x_0)) = \inf_{n\geq 1} (\sup_{i\geq n} f_i(x_0))$$

and it applies to every x_0 so

$$\lim_{n\to\infty} sup\ f_n(x) = \lim_{n\to\infty} (\sup_{i\geq n} f_i(x)) = \inf_{n\geq 1} (\sup_{i\geq n} f_i(x))$$

The same logic can be used on $lim\ inf$

$$\lim_{n\to\infty} inf f_n(x) = \lim_{n\to\infty} (\inf_{i\geq n} f_i(x)) = \sup_{n\geq 1} (\inf_{i\geq n} f_i(x))$$

Just like set angle, if $lim\ sup_{n\to\infty}\ f_n(x) = lim\ inf_{n\to\infty} f_n(x)$, then we call $f_n(x)$ has limit, it is written as

$$\lim_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x) = f(x)$$

Comparing the set angle and function angle, we can find these two can derive the same results. The only difference comes with $f_i(x_0)$ or whether an element is in set A , however the latter can be seen as a special of the former

$$f(\omega) = \begin{cases} 0 & \omega \notin A \\ 1 & \omega \in A \end{cases}$$

In words, set can be seen as a special function of which all the functions values are either 1 or 0

So go beyond the subject we are talking about, any action we take on function should be able to be applied to set naturally.

1.2. Semi-algebras, Algebras, and σ algebras

1.2.1. Semialgebra

A collection S of sets is said to be a semialgebra

$$(1)\ if\ S, T \in \mathcal{S},\ then\ S \cap T \in \mathcal{S}$$

$$(2)\ if\ S \in \mathcal{S},\ then\ S^c\ is\ finite\ disjoint\ union\ of\ sets\ in\ \mathcal{S}$$
$$i.e.\ S^c = T_1 \cup \ldots \cup T_n,\ and\ T_i s \in \mathcal{S}\ are\ finite\ and\ disjoint$$

$$(3)\ \Omega \in \mathcal{S}$$

Notice that in the second point, the definition doesn't require S^c to be in $\mathcal{S}$

1.2.2. Algebra

From Shijian Yan's book, we can see there are so many definitions of algebra. Here we list two of them.

A collection $\mathcal{A}$ of subsets of Ω is called an algebra (or field) if

$$\mathcal{A}\ is\ closed\ under\ complement\ and\ finite\ intersection$$

$$(1)\ if\ A, B \in \mathcal{A},\ then\ A \cap B \in \mathcal{A}$$

$$(2)\ if\ A \in \mathcal{A},\ then\ A^c \in \mathcal{A}$$

$$(3)\ \Omega \in \mathcal{S}$$

or

$$\mathcal{A}\ is\ closed\ under\ complement\ and\ finite\ union$$

$$(1)\ if\ A, B \in \mathcal{A},\ then\ A \cup B \in \mathcal{A}$$

$$(2)\ if\ A \in \mathcal{A},\ then\ A^c \in \mathcal{A}$$

$$(3)\ \Omega \in \mathcal{S}$$

These two definitions actually can let us see the "bridge" effect algebra has between semialgebra and σ algebra.

The first definition resembles the one of semialgebra, comparing the two, we find that when a collection of sets satisfy the second point of the algebra definition, it will automatically satisfy the definition of semialgebra, in other word, **an algebra is a semi-algebra**

1.2.3. σ algebra

Then let's compare the second definition and the definition of σ algebra, but first we have to adress the meaning of finite. finite corresponds to the first point, because when we pick A and B out of $\mathcal{A}$, we know for sure that $A \cup B \in \mathcal{A}$, then we can pick C out of $\mathcal{A}$, we then know $(A \cup B) \cup C \in \mathcal{A}$ as well. The step can go on and on. However since the first point doesn't have the infinity symbol meaning that it doesn't specify the bahavior when reaching the infinity, so as long as the number of sets that we pick out $\mathcal{A}$ is smaller than infinity, union is closed, in another word, $\mathcal{A}$ is closed under finite union.

$$\mathcal{F}\ is\ closed\ under\ complement\ and\ countable\ union$$

$$(1)\ if\ A_i \in \mathcal{F}\ is\ a\ countable\ sequence\ of\ sets,\ then\ \cup_i A_i \in \mathcal{F}$$

$$(2)\ if\ A \in \mathcal{F},\ then\ A^c \in \mathcal{F}$$

$$(3)\ \Omega \in \mathcal{F}$$

we can see that the only difference between an algebra and a σ algebra is that σ algebra pushes union operation to infinity, which means whether the number of sets we pick out of $\mathcal{F}$ is finite or countably infinite, union operation is closed. So when a collection of sets is σ algebra, it is an algebra. In other words, **a σ algebra is an algebra**

This is exactly the reason why it's called σ algebra, σ means countable sum, countable union. Whenever you see σ like σ finite, or σ additive, it must push the sum or union operation from finite to countably infinite.

The pair $(\Omega, \mathcal{F})$ is called a measurable space. The sets of $\mathcal{F}$ are called measurable sets. In other words, a set is measurable only when it is in a σ algebra.

The most important thing of σ algebra for probability is not it's definition, but it's relationship with information.

Information is something accurate about a system, it is essentially the final observable result meaning when we have more information, we are getting close to the final observable result.

As an individual, I want to study a stochastic system. Before I really dive into it, I only have one piece of information: all the possible outcomes, or the description of all the possible outcomes. For example, if I want to study coin toss of finite or maybe countbly infinite times (here we study 3 times), I will know the outcome would be a list of H or T. We call it Ω , each possible outcome ω is an element in this set.

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

A random experiment is performed, i.e., an element ω_0 of Ω is selected. The value of ω_0 is partially but not fully revealed to us, which means we only have partial information about the result.

How do we get the partial information? We pray to the system: if you don't want tell me the result, that's fine but allow me to construct a σ algebra, at least you could tell me which of these sets ω_0 belongs to. So we receive the information "ω_0 locates in A" and it holds true for every possible ω in Ω . Essentially the statement becomes : σ algebra guarantees us to know which sets ω belongs to, once the experiment is performed.

So the σ albegra's structure is designed by us before the information and we are confident that the system will satisfy us by tell us which set ω belongs to. That is why σ albegra represents information. ($f(x)$ will tell us the value of $f(x_0)$ but won't tell us x_0 , and we design the structure of $f(x)$)

When we don't want to receive partial information, the division would be:

$$\mathcal{F}_0 = \{\emptyset, \Omega\}$$

When the experiment is perfomed, of course ω will be in Ω and not $\emptyset$, and I don't need any partial information for this conclusion. Then I wonder : I want to know the information of the firs coin toss, so I design $\mathcal{F}_0$ as

$$A_H = \{HHH, HHT, HTH, HTT\}$$

$$A_T = \{THH, THT, TTH, TTT\}$$

$$\mathcal{F}_1 = \{\emptyset, A_H, A_T, \Omega, \ all\ unions\}$$

When the experiment is perfomed and I am told ω is in A_H , then I'll know "the first coin toss is head". After $\mathcal{F}_1$ we wonder : I want to know both the first two coin tosses, then the division would be:

$A_{HH} = \{HHH, HHT\}$

$A_{HT} = \{HTH, HTT\}$

$A_{TH} = \{THH, THT\}$

$A_{TH} = \{TTH, TTT\}$

$\mathcal{F}_2 = \{\emptyset, A_{HH}, A_{HT}, A_{TH}, A_{TT}, \Omega, \ and\ all\ sets\ which\ can\ be\ built\ by\ taking\ unions\ of\ the$

So σ algebra is our plan to receive partial information of the result which is guarantees to be realized

---

To sum up:

- **An algebra is a semi-algebra**
- **A σ algebra is an algebra**

Since the definition of algebra is diverse, we'll make a choice based on the particular circumstance we are studying.

1.3. Generated classes (Minimal classes)

The way to generate one from another is stated as following:

If $\mathcal{S}$ is a semialgebra, then $\overline{\mathcal{S}} = \{finite\ disjoint\ unions\ of\ sets\ in\ \mathcal{S}\}$ is an algebra, called the algebra generated by $\mathcal{S}$, the meaning is that using this act, we can construct a unique algebra from a semialgebra.

Using a random collection of sets $\mathcal{A}$, we can construct many σ algebra, but we focus on the smallest one, and call it the σ algebra generated by $\mathcal{A}$. So based on $\overline{\mathcal{S}}$ and $\mathcal{S}$, we can generated $\sigma(\overline{\mathcal{S}})$ and $\sigma(\mathcal{S})$. We have

$$\mathcal{S} \subset \overline{\mathcal{S}} \subset \sigma(\overline{\mathcal{S}}) = \sigma(\mathcal{S})$$

1.4. Monotone class (m-class), π class, and λ class

These three classes are here to help us to check σ algebra

When we choose a sequence of sets $A_1, A_2 \ldots$ out of Ω , two types of limit would be

$$Infinitely\ often: \quad \overline{\lim_{n\to\infty}} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

$$Ultimately: \quad \underline{\lim_{n\to\infty}} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

and when the sequence is monotone sequence, the above two limit will converge to one, and we call this sequence has limit

$$\overline{\lim_{n\to\infty}} A_n = \underline{\lim_{n\to\infty}} A_n = \lim_{n\to\infty} A_n$$

Based on whether the sequence is increase or decrease, the limit has two forms

$$A_1 \subset A_2 \subset \ldots \Rightarrow \lim_{n\to\infty} A_n = \cup_i^{\infty} A_i$$

$$A_1 \supset A_2 \supset \ldots \Rightarrow \lim_{n\to\infty} A_n = \cap_i^{\infty} A_i$$

Monotone class: we choose any monotone(increase or decrease) sequence from the class, its limit would still be in the class.

π **class** : closed under finite intersection

λ **class** : too long, read the book

Book listed the relation between these three and σ algebra

Chapter 2 Measure Theory

2.1. Definition of Measure

The two definitions related to set function finite:

- Finite: $\forall A \in \mathcal{A}, |\mu(A)| < \infty$, it means that the set function value of every element in $\mathcal{A}$ is finite, then we can call the set function is finite.
- σ finite (countable): $\exists \{A_i\} \in \mathcal{A}, \Omega = \cup_{i=1} A_i \ and \ \mu(A_i) < \infty$, it means that we can find a countable (finite or countable infinite) sequence in $\mathcal{A}$, their union can form Ω ， and everyone's set function value is finite.

compared to Finite which is easy to understand, when it comes to the set function value, σ finite doesn't requires all the elements in class are finite, it only requires we can at least find one sequence of sets, and these sets' set function values are finite. Since we drop the standard from one aspect, we have to raise the standard from another: the sequence of sets that we find have to comprise Ω

Another angle is: when defining σ finite, we don't just define the property of μ , we will add some restriction to the class as well. In other words, σ finite is not just the property of set function, it is the property of both set function and class.

A measure being σ finite is a weaker condition than being finite, i.e. all finite measures are σ finite but there are σ finite measures that are not finite.

Besides finite, the reason why we need σ finite is that in some circumstances where we restrict measure to be σ finite, we can get some really good property

The two definitions related to set function additivity: for pairwise disjoint sets

- Additive (finite) : $\mu(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$
- σ additive or countably additive (countable) : $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Every σ additive function is additive but not vice versa

What's the relationship between finite and additivity? They are independent to some extent, meaning one is true doesn't guarantee the other stands true. The former describes the value of set function and the latter describes their behavior when being added up.

In section 2.5, author shows that a measure can be neither finite nor σ finite, again it proves that finite and additivity are independent to some extent

After the above discussion, we can talk about measure. Measure is set funtion with two restrictions

- The set function value of every element in class is nonnegative
- The set function is σ additive, meaning no matter how we choose or how many disjoint elements we choose from class(finite or countably infinite), the set function value of their union equals to the sum of their set function values

Any set function that satisfies the above two can be called Measure.

So far we have not put any restriction on the space Ω or the class $\mathcal{A}$, meaning semialgebra, algebra, σ algebra... they all can have measure.

But the following nouns only stands when the class is σ algebra. If $\mathcal{F}$ is σ algebra and μ is a measure

- $(\Omega, \mathcal{F})$ is called measurable space
- The sets of $\mathcal{F}$ are called measurable sets or $\mathcal{A}$ measurable
- $(\Omega, \mathcal{F}, \mu)$ is called measure space
- If $\mu(\Omega) = 1$, $(\Omega, \mathcal{F}, \mu)$ is called probability space

After the definition of additivity, we can talk about continuity of measure(keep in mind the class we discuss below is stiil all kinds of classes, not just σ algebra).

Since set doesn't have size like number, but continuity means that we have to approach a set somehow, so we use monotone set of sequence to approach a set from above and below

- Measure is always continuous from below
- Measure is continuous from above conditionally on the assumption that $\mu(A_m) < \infty$ for some finite m
- Finite measures (such as probability measure) are always continuous.

2.2. Properties of measure

2.1 only gave us the basic properties of measure meaning the commom properties measure has based on its definition when the class is of any kind. And the basic property is continuity.

So 2.2 consider the properties of measure when the class is specific: semialgebra, algebra, σ algebra

2.2.1. Case I: semialgebras

Revisit its definition

$$(1)\ if\ S, T \in \mathcal{S},\ then\ S \cap T \in \mathcal{S}$$

$$(2)\ if\ S \in \mathcal{S},\ then\ S^c\ is\ finite\ disjoint\ union\ of\ sets\ in\ \mathcal{S}$$
$$i.e.\ S^c = T_1 \cup \ldots \cup T_n,\ and\ T_i s \in \mathcal{S}\ are\ finite\ and\ disjoint$$

$$(3)\ \Omega \in \mathcal{S}$$

When μ is a nonnegative additive set function on a semialgebra $\mathcal{S}$, we have

$$(1)\ Monotonicity$$

$$(2)\ \sigma - Subadditivity$$

Keep in mind, some of the properties in here only requires additive set function instead of measure.

The proof of first point uses semialgebra's definition.

2.2.2. Case II: algebras

Beside all the properties it inherits from semialgebra, it has its own σ subadditivity. It directly uses measure here

2.2.3. Case III: σ algebras

2.3. Probability measure

A bunch of properties which can be useful.

Based on the definition of measure on σ algebra, it adds two more restrictions

- $\forall A \; in \; \mathcal{F}, \mu(A) \leq 1$
- $\mu(\Omega) = 1$

2.4. Some examples of measure

In this section, I want to talk about discrete probability measure

Revisit the definition of probability measure

$$(1) \; (\Omega, \mathcal{F}, P) \; is \; measure \; space$$

$$(2) \; P(A) \leq 1 \; for \; any \; A \in \mathcal{F}$$

$$(3) \; P(\Omega) = 1$$

And the definition of discrete probability measure

(1) $(\Omega, \mathcal{F}, P)$ *is measure space and* Ω *is either finite or countably infinite*

(2) $P(A) \leq 1$ *for any* $A \in \mathcal{F}$

(3) $P(\Omega) = 1$

So the only difference is that the measure space is either finite or countably infinite which means we can count the element in set. When we calculate the probability of a set in general probability measure, the expression is

$$P(A) = P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

This is directly from the definition of measure: the measure of the union of disjoint sets is the sum of the measures of all the sets. The reason why we have to cut A into little pieces A_i is that we don't know whether or not we can count. However when the space is countable, we can actually count the elements in the set

$$P(A) = P(\{\omega_1, \omega_1, \dots\}) = P(\sum_{i=1}\{w_i\}) = \sum_{i=1} P(\{w_i\}) = \sum_{i=1} P(w_i) = \sum_{w \in A} P(\omega)$$

2.5. Extension of set functions (or measures) from semialgebras to algebras

In chapter 1, after we defined all the classes that we need, we discuss the relationship in classes: does one class belong to another class? how do we create one class from another class? In other words, the dynamic property of classes.

In chapter 2, the logic is similar, we have discussed the meaning of measure, and its properties on different classes. Now it is time to ask ourselves, how do measures on different classes related to each other? In other words, the dynamic property of measure based on the dynamic shift of classes.

This dynamic property has a name: extension

Definition is: $\mathcal{A}$ *and* $\mathcal{B}$ *are two classes of subsets of* Ω *with* $\mathcal{A} \subset \mathcal{B}$ *,* μ, ν *are two set functions (measures) defined on* $\mathcal{A}, \mathcal{B}$ *, respectively such that*

$$\mu(A) = \nu(A), \forall A \in \mathcal{A}$$

then ν *is said to be an extension of* μ *from* $\mathcal{A}$ *to* $\mathcal{B}$*, and* μ *the restriction from* $\mathcal{B}$ *to* $\mathcal{A}$

Note that the above definition only involves two classes that has the relationship $\mathcal{A} \subset \mathcal{B}$, meaning two classes of any kind can form extension from one to another. The reason we mainly focus on semialgebra, algebra and σ algebra is that the extension on these classes can help us study probability.

In this section, we address a simplified version of the question: the extension of measure from semialgebra to algebra

1. *Let μ be a non-negative additive set function (or measure) on a semialgebra $\mathcal{S}$, then μ has a unique extension $\overline{\mu}$ to $\overline{\mathcal{S}} = \mathcal{A}(\mathcal{S})$, such that $\overline{\mu}$ is additive*
2. *Moreover, if μ is σ additive on S (which implies that μ is a measure on $\mathcal{S}$), then so is $\overline{\mu}$ on $\overline{\mathcal{S}}$*

2.6. Outer measure

μ is a measure on semialgebra $\mathcal{S}$, for any $A \subset \Omega$

$$\mu^*(A) = inf\ \{\sum_{i=1} \mu(A_i) | A \subset \cup_{i=1} A_i, A_i \in \mathcal{S}\}$$

- The logic is we pick a subset out of Ω which is not necessary an element in $\mathcal{S}$
- We use it to pick a squence(countable) of sets out of $\mathcal{S}$, and the only requirement is the union of the sets in the squence has to comprise A , of course we can find many sequences that can do this.
- We then calculate the sum of measure values of the sets in one sequence, each sequence corresponds to one value
- We compare these values and pick the smallest and call it the outer measure of A , and μ^* (can be understood as a function or the above four steps) is the outer measure induced by μ

We can see the purpose of outer measure is formally pushing measure from semialgebra to power set, however $\mu^*(A)$ may not satisfy the definition of measure. But still it has some properties similar to measure

$$(1)\ \forall A \in \mathcal{S},\ \mu^*(A) = \mu(A)$$

$$(2)\ Monotonicity$$

$$(3)\ \sigma - Subadditivity$$

The thing we are about discuss is reverse version of some of the things that we discussed earlier, by which I mean before this section we use set theory as the corner stone for measure theory. We have to define all kinds of classes and then define the measure on them, the measure itself doesn't hold any restrictions on classes, only have different properties on different classes.

But now we intend to screen all the sets in power set using measure. A set $A \subset \Omega$ is said to measurable w.r.t. an outer measure μ^* iff (if and only if) for any $D \subset \Omega$, it has

$$\mu^*(D) \geq \mu^*(A \cap D) + \mu^*(A^c \cap D)$$

When μ^* is fixed, we gather all the subsets of Ω that satisfy the above inequality, and call the class $\mathcal{A}^*$. No surprise, $\mathcal{A}^*$ would have many fantastic properties:

(1) $\mathcal{A}^*$ *is a* σ *algebra*

(2) $if\ A = \sum_{i=1} A_i,\ with\ A_i \subset \mathcal{A}^*\ then\ for\ any\ B \subset \Omega,\ \ \mu^*(A \cap B) = \sum_{i=1} \mu^*(A_i \cap B)$

(3) $(\Omega, \mathcal{A}^*, \mu^*|_{\mathcal{A}^*})$ *is a measure space. Furthermore* $\mu^*|_{\mathcal{A}^*}$ *is an extension of* μ *from* $\mathcal{S}$ *t*

2.7. Extension of measures from semialgebras to σ algebras

In chapter 1, we generate a σ algebra $\sigma(\mathcal{S})$, and in above section we generate σ algebra $\mathcal{A}^*$, both are from $\mathcal{S}$. What is the difference between the two?

$$S \subset \overline{S} \subset \sigma(\mathcal{S}) \subset \mathcal{A}^* \subset \mathcal{P}(\Omega)$$

We can see that the σ algebra $\mathcal{A}^*$ generated by the outer measure is bigger that the σ algebra $\sigma(\mathcal{S})$ generated directly by S

From there, we have the following theorem(Caratheodory Extension Theorem)

Let μ *be a measure on a semialgebra* S

1. μ *has an extension to* $\sigma(S)$ *, denoted by* $\mu|_{\sigma(S)}$ *, so* $(\Omega, \sigma(S), \mu|_{\sigma(S)})$ *is a measure space*
2. *Furthermore,* $\mu|_{\sigma(S)} = \mu^*|_{\sigma(S)}$ *, i.e. this extension can be simply taken to be the restriction of measure* $\mu^*|_{\mathcal{A}^*}$ *to* $\sigma(S)$
3. *If* μ *is* σ *finite, then the extension in 1 is unique*

We know from chapter 1, using S to generate $\sigma(S)$ is possible (actually any class can generate a smallest σ algebra) , here in the first point, what the author is trying to say is that not only can we generate a smallest σ algebra, we can also extend the measure on S to the smallest σ algebra that we generated.

The second point, if we just mention the outer measure without any purpose, it's just stupid. Here the author states its purpose: to construct measure of σ algebra $\sigma(\mathcal{S})$. If we have already known σ algebra $\mathcal{A}^*$ and its measure $\mu^*|_{\mathcal{A}^*}$, and since $\sigma(\mathcal{S}) \subset \mathcal{A}^*$, it is pretty straightforward to keep the measure and shrink the class.

The third point states that if μ is σ finite, all the shenanigans we did in the first and second point will get us the exact same measure.

Then we wonder how can all these relate to probability space? First we can have (Ω, S, μ) with $\mu(\Omega) = 1$. Check out the definition of σ finite, we can find that when $\mu(\Omega) = 1$, measure will satisfy σ finite with no doubt. Then it can extend uniquely to $\sigma(S)$, thus we have $(\Omega, \sigma(S), \mu^*)$ with $\mu^*(A) = \mu(A), A \in S$ and of course $\mu^*(\Omega) = 1$, so it is a probability space.

2.8. Completion of a measure

Definition: Let $(\Omega, \mathcal{F}, \mu)$ be measure space and $N \subset \Omega$

1. *N is a μ - null set iff $\exists B \in \mathcal{F}$ with $\mu(B) = 0$ such that $N \subset B$*
2. *$(\Omega, \mathcal{F}, \mu)$ is a complete measure space if every μ - null set $N \in \mathcal{F}$*

Once again, I have to state that measure space means the class is σ algebra and μ is measure. So logic of the definition is

1. We pick out a subset N out of Ω
2. Then we pick one or more elements the measure of which is 0 from $\mathcal{F}$
3. We then compare N with these elements(again the size of which maybe one maybe more than one) , and just so luckily we can find at least one element that covers N , and we can call N a μ - null set
4. We gather all N , we can actually use the elements we found in step 2 and all their subsets are all the N
5. We then see whether all the N are in $\mathcal{F}$, if they are in, we call $(\Omega, \mathcal{F}, \mu)$ complete measure space

From the definition, we can derive the fact: if N is only a μ - null set, it is not nacessary in $\mathcal{F}$, in other words, not $\mathcal{F}$ measurable, however if $(\Omega, \mathcal{F}, \mu)$ is a complete measure space , N is defintely in $\mathcal{F}$.

The premise of last sentence is fixing the space and studying its structural property. What if the space can change? meaning if it is not complete, can we add something to make it complete? It turns out we can

Given a measure space $(\Omega, \mathcal{F}, \mu)$, there exists a complete measure space $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$ such that $\mathcal{F} \subset \overline{\mathcal{F}}$ and $\overline{\mu} = \mu$ on $\mathcal{F}$, we call $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$ the completion of $(\Omega, \mathcal{F}, \mu)$

One thing I need to mention again is that the extension of measure can happen to any two classes as long as one is the subset of the other. Here what we prensent is a special case where both classes are σ algebra

How do completion relate to other sections?

Theorem: Let μ be a σ finite measure on a semialgebra S , μ^* be the outer measure induced by μ , and $\mathcal{A}^*$ the σ algebra consists of all the μ^* measurable sets. Then $(\Omega, \mathcal{A}^*, \mu^*|_{\mathcal{A}^*})$ is the completion of $(\Omega, \sigma(S), \mu^*|_{\sigma(S)})$

One way to extend measure from semialgebra to σ algebra $\sigma(S)$ in last section is to first construct $\mathcal{A}^*$ and then shrink it to $\sigma(S)$, from which we know how to go from $(\Omega, \mathcal{A}^*, \mu^*|_{\mathcal{A}^*})$ to $(\Omega, \sigma(S), \mu^*|_{\sigma(S)})$.

What this section has taught us is that how to go from $(\Omega, \sigma(S), \mu^*|_{\sigma(S)})$ to $(\Omega, \mathcal{A}^*, \mu^*|_{\mathcal{A}^*})$ with a minor restriction: μ is σ finite which appeared in last section as well.

When proving the above theorem, author proves that the gap between $\mathcal{A}^*$ and $\sigma(S)$ is all μ_σ null sets (even without the restriction of σ finite)

$$\mathcal{A}^* = \sigma(S) + \{all\ \mu_\sigma - null\ set\}$$

2.9. Construction of measures on a σ algebra

The general idea is to firstly constuct a semialgebra and a measure, then extend them to $\mathcal{A}^*$ and μ^* , then shrink to $\sigma(S)$

The first issue is how to find a measure for a semialgebra, we have the following theorem

Theorem: Let μ be a nonnegative set function a semialgebra S . For pairwise disjoint sets, If

1. μ is additive on S (finite)
    - i.e. $\mu(A) = \sum_1^n \mu(A_i)$, whenever $A_n \in S$ ans $A = \sum_1^n A \in S$
2. μ is σ subadditive on S (countable)
    - i.e. $\mu(A) \leq \sum_{i=1} \mu(A_i)$, whenever $A, A_i \in S$ and $A \subset \cup_{i=1} A_i$

Then μ is a measure on S

Let's compare the theorem with theorem from section 2.2.1 (written below), the properties of measure when we know for sure there is a measure on semialgebra

Theorem: When μ is a nonnegative additive set function on a semialgebra $\mathcal{S}$, we have

- Monotonicity: $A \subset B \Rightarrow \mu(A) \leq \mu(B)$
- σ Subadditivity:
    a. $\sum_{i=1} A_i \subset A \Rightarrow \sum_{i=1} \mu(A_i) \leq \mu(A)$

b. Further assume that μ is σ additive (hence a measure) , then
$A \subset \sum_{i=1} A_i \Rightarrow \mu(A) \leq \sum_{i=1} \mu(A_i)$

The logic of these two theorem is inherently different, the former describes what it takes to become a measure, and the latter describes the properties set function should have if it becomes a measure. Naturally the former description is much shorter because it only takes out a few things from the latter theorem.

2.9.1. Lebesgue and Lebesgue-Stieltjes measures

Based on Borel measurable space which first appeared in section 1.6, we construct L-S measure

The 6th remark of this section tell us why we should have all the trouble in the world to construct σ algebra, it's because the power set is just too large.

2.9.2. Relationship between probability measures and distribution functions

Distribution function is considered as a special case of F in L-S measure, meaning the foundation of distribution function is $(\mathcal{R}, \mathcal{B})$. When we use d.f. on $(\mathcal{R}, \mathcal{B})$, it will automatically becomes probability space $(\mathcal{R}, \mathcal{B}, P)$

Definition:

A real-valued function F on $\mathcal{R}$ is distribution function (d.f.) if

- $F(-\infty) = \lim_{x \to -\infty} F(x) = 0,\ F(\infty) = \lim_{x \to \infty} F(x) = 1$
- F is nondecreasing, i.e., $F(x) \leq F(y)\ if\ x \leq y$
- F is right continuous, i.e., $F(y) \downarrow F(x)\ if\ y \downarrow x$

The relation of distribution function and probability measure on $(\mathcal{R}, \mathcal{B})$

$$F(x) = P((-\infty, x]),\ x \in \mathcal{R}$$

2.9.3. Decomposition of distribution functions

After the introduction of distribution function, this and the next section talks about the properties a distribution function must have

2.10. Radon-Nikodym theorem

Definition: μ and ν are two measures on the measurable space $(\Omega, \mathcal{F})$, if $\forall A \in \mathcal{F}\ and\ \mu(A) = 0$, we have $\nu(A) = 0$, we say ν is absolutely continuous w.r.t. μ , written as $\nu \ll \mu$

Theorem: Given a measurable space $(\Omega, \mathcal{F})$, if a measure ν on $(\Omega, \mathcal{F})$ is absolutely continuous with respect to a sigma-finite measure μ on $(\Omega, \mathcal{F})$, then then there is a measurable function f on Ω and taking values in $[0, \infty)$, such that $\nu(A) = \int_A f d\mu$ for any measurable set A

The theorem is too abstract, let's use an example to understand it

When we calculate the probability of a certain event from pdf, we write $P(A) = \int_A f(x)dx$, the logic is that first we have to assign probability density function value to all the

- μ is lebesgue measure which means $\mu((a, b]) = b - a$
- ν is continuous probability distribution which implies $P(\Omega) = \nu(\Omega) = 1$
- The theorem is to say we can always find a probability density function such that $P(A) = \int_A f d\mu$

2.11. Change of Measure

Definition : In $(\Omega, \mathcal{F}, P)$, Z is a almost everywhere nonnegative random variable, and $E(Z) = 1$. Define $\tilde{P}(A) = \int_A Z(\omega)dP(\omega),\ \forall A \in \mathcal{F}$, we can proof $\tilde{P}(A)$ is a probability measure

What change of measure does is shrinking the probability of some sets and magnify the probability of other sets. One property of change of measure is if X is a nonnegative random variable, we then have $\tilde{E}(X) = E(XZ)$

Example : $\Omega = [0, 1]$, P is uniformly distributed and we let $Z(\omega) = 2\omega$. First we check the random variable: it is nonnegative and its expectation

$$EZ = \int_0^1 2\omega dP(\omega) = \int_0^1 2\omega d\omega = 1$$

so it satisfy the definition. The we can deduce the new probability measure

$$\tilde{P}([a, b]) = \int_a^b 2\omega d\omega = b^2 - a^2$$

We also can see

$$d\tilde{P} = 2\omega dP \Rightarrow Z = \frac{d\tilde{P}}{dP}$$

Z is Radon-Nikodym derivative.

Chapter 3 Random Variables

3.1. Mappings

Definition:

Let $X : \Omega_1 \to \Omega_2$ be a mapping

For every subset $B \in \Omega_2$, the inverse image of B is

$$X^{-1}(B) = \{\omega : \omega \in \Omega_1, X(\omega) \in B\} = \{X \in B\}$$

For every class $\mathcal{G} \subset \Omega_2$, the inverse image of $\mathcal{G}$ is

$$X^{-1}(\mathcal{G}) = \{X^{-1}(B) : B \in \mathcal{G}\}$$

Why does the author have to introduce X^{-1} ? It is because we will use inverse image frequently and it is not inverse function we learned in high shool. The existence of inverse funtion requires one element from Ω_1 and one element from Ω_2 form a unique pair, meaning we can find one using the other.

However mapping here only requires we project elements from Ω_1 to Ω_2 meaning one element from Ω_2 can be the projection of multiple elements from Ω_1 , so by definition it doesn't have inverse function. When we want to locate elements from Ω_1 , we can't use inverse function, the author uses inverse image which is to gather all the elements that project onto one element from Ω_2

Let's look at an example: $(\Omega_1, \mathcal{A}) \to X \to (\Omega_2, \mathcal{B})$

$$A = \{\omega_1, \omega_2, \omega_3\} \in \mathcal{A}$$

$$X = I_A(\omega) = \begin{matrix} 1 & if\ \omega \in A \\ 0 & if\ \omega \notin A \end{matrix}$$

When we use the inverse image

$$\{I_A \in B\} = \begin{matrix} \emptyset & if\ 0\ \notin B,\ 1 \notin B \\ A^c & if\ 0\ \in B,\ 1 \notin B \\ A & if\ 0\ \notin B,\ 1 \in B \\ \Omega & if\ 0\ \in B,\ 1 \in B \end{matrix}$$

When we use the inverse image on class level, we typically only get a small portion of what we had originally. Our function X maps $\omega_1, \omega_2, \omega_3$ onto 1 , then we use 1 to gather all the elements to form a set, it would be $\{\omega_1, \omega_2, \omega_3\}$. So we lose $\{\omega_1, \omega_2\}$, $\{\omega_1, \omega_3\}$, $\{\omega_2, \omega_3\}$, $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$ which are usually parts of $\mathcal{A}$. To put it into plain words, X^{-1} usually has shrinking effect on class level.

It is worth mentioning, however, if the class $\mathcal{B}$ is a σ algebra, $X^{-1}(\mathcal{B})$ would be a σ algebra as well, in other words, the result $X^{-1}(\mathcal{B})$ that have shrinked compared with $\mathcal{A}$ is a σ algebra, as long as $\mathcal{B}$ is a σ algebra, and it has nothing to do with the status of $\mathcal{A}$

To sum up

- The mapping result can be a subset of Ω_2
- $X^{-1}(\mathcal{B})$ is usually smaller than $\mathcal{A}$
- If $\mathcal{B}$ is a σ algebra, $X^{-1}(\mathcal{B})$ is a σ algebra

3.2. Measurable mapping

Definition:

1. $(\Omega_1, \mathcal{A})$ and $(\Omega_2, \mathcal{B})$ are measurable spaces ($\mathcal{A}, \mathcal{B}$ are σ algebra). $X : \Omega_1 \to \Omega_2$ is measurable mapping if $X^{-1}(B) \in \mathcal{A}, \forall B \in \mathcal{B}$
2. X is a measurable function if $(\Omega_2, \mathcal{B}) = (\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$
3. X is a Borel (measurable) function if $(\Omega_1, \mathcal{A}) = (\mathcal{R}^m, \mathcal{B}(\mathcal{R}^m))$ and $(\Omega_2, \mathcal{B}) = (\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$

The difference between mapping and measurable mapping is that the former is only defined on space, and the latter is defined on measurable space (space + σ algebra)

So the logic of definition's first point is that we pick any set B from $\mathcal{B}$, using the inverse image of X , we can find $X^{-1}(B)$ is always in $\mathcal{A}$. It implies that when $\mathcal{A}$ is bigger than or equal to $\mathcal{B}$ at the view of X , we call X a measurable mapping.

"measurable" comes from measurable space, "mapping" comes from X .

The second point is pretty natural, by the definition of function, a mapping becomes a function only when there are numbers involved. When we switch Ω_2 to $\mathcal{R}^n$, the result of X becomes intervals of number, so mapping becomes function.

The third point, Borel means all the things related to set become numbers

So we brought up the definition of measurable mapping, but the definition is not practical because if we want to check a mapping, we just can't list out all the elements in $\mathcal{B}$, so for simplicity, we turn to the next theorem

Theorem:

$X : (\Omega_1, \mathcal{A}) \to (\Omega_2, \mathcal{B})$ is a measurable mapping if $\mathcal{B} = \sigma(\mathcal{C})\ and\ X^{-1}(C) \in A, \forall C \in \mathcal{C}$

How does it simplify checking procedure? Before the theorem, we have to check all the elements in $\mathcal{B}$, but here we shrink it to $\mathcal{C}$, all we have to check is all its subsets which are considerably smaller than the size of $\mathcal{B}$

3.3. Random Variables (Vectors)

Derived from the second point of last definition, we have the definition of random variable

Definition:

A random variable X is a measurable function from $(\Omega,\mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$, i.e.
$X^{-1}(B) \in A, \forall Borel\ set\ B \in \mathcal{B}$

Notation1: $X^{-1}(B)$ can be denoted as $\{X \in B\}$, just like using $f(x) < y_0$ to denote a certain interval for x . So remember whenever you see something like $\{X < x\}$, it represents the set in $\mathcal{A}$ that make $X(set\ on\ \mathcal{A})$ become $(-\infty, x)$

Notation2: $X \in \mathcal{A}$ means X is $\mathcal{A}$ measurable

Here I have to discuss the essence of random variable : a function X is essentially a σ algebra $\sigma(X)$ with the fact that each atom come with a serial number. "X is $\mathcal{F}$ measurable" means that the structure of $\sigma(X)$ is coverd by $\mathcal{F}$. Two other ways to explain it

- $\mathcal{F}$ will contain more information than X
- We divide the sets in $\sigma(X)$ into little pieces and add them back to $\sigma(X)$, we get $\mathcal{F}$

We can see that random variable's base is only measurable space, it doesn't necessarily relate to probability. However some people define r.v. only on probability space. Using definition to check r.v. is just exhausting, and we can simplify the procedure using the theorem we mentioned in the last section.

Let's revisit the definition of Borel σ algebra: The smallest σ algebra generated by the collection of all finite open intervals on the real line $R = (-\infty, \infty)$ is called the Borel σ algebra, denoted by $\mathcal{B}$. The elements of $\mathcal{B}$ are called Borel sets. The pair $(\mathcal{R}, \mathcal{B})$ is called the (1-dimensional) Borel measurable space.

$$\mathcal{B} = \sigma(all\ finite\ open\ intervals\ on\ the\ real\ line)$$

But it is not the only way to generate Borel σ algebra

$$\mathcal{B} = \sigma(\{[-\infty, x] : \forall x \in \mathcal{R}\})$$
$$or$$
$$\mathcal{B} = \sigma(\{[-\infty, x] : \forall x \in \mathcal{D}\})$$

$\mathcal{D}$ is a dense subset of $\mathcal{R}$, A subset $\mathcal{D}$ of $\mathcal{R}$ is said to be a "dense" subset if for every $\omega \in R$, there is a sequence $\{x_n\}$ of numbers in $\mathcal{D}$ which converges to ω , e.g.
$\mathcal{D} = all\ rational\ numbers$

From the theorem of last section, we don't have to check all the sets in $\mathcal{B}$, and we can check the class that generated $\mathcal{B}$ instead. So the checking procedure becomes:

$$\{X \leq x\} = X^{-1}([-\infty, x]) \in \mathcal{A}, \forall x \in \mathcal{R}$$
$$or$$
$$\{X \leq x\} = X^{-1}([-\infty, x]) \in \mathcal{A}, \forall x \in \mathcal{D}$$

When we use this procedure, we have to remind ourselves $[-\infty, x]$ is just a part of $\mathcal{B}$, meaning $\{X \leq x\}$ is just a part of the sets that we should check by definition of r.v.

3.4. Construction of random variables

3.4.1. Algebraic operations

We only have one $(\Omega, \mathcal{A})$ and one $(\mathcal{R}, \mathcal{B})$, one function that can map the former to the latter is what we have described above. Here we consider two functions, each doing its own mapping

$$X(A) = B_1\ and\ Y(A) = B_2$$

If we combine the two functions, what would the result be?

$$Z(A) = Z_{X,Y}(A) = ?$$

Author states that, if
$Z = aX + bY, max\{X, Y\}, min\{X, Y\}, X^2, XY, X/Y\ (Y(\omega) \neq 0), X^+, X^-, |X|$,
then Z is a random variable

So the essence of first part of the section is construction of r.v. using finite functions

3.4.2. Limiting operations

Then once again we need to consider countably infinite, the "countably infinite" we discuss here is of a function sequence

Theorem:

$X_1, X_2, \ldots$ are r.v. on $(\Omega, \mathcal{A})$ (i.e. $X_i \in \mathcal{A}$) , the functions we list below are r.v.

- $sup_i X_i$, $inf_i X_i$, $lim\ sup_i X_i$, $lim\ inf_i X_i$
- If $X(\omega) = \lim_n X_n(\omega)$ for every ω , then X is a r.v. (i.e., $X \in \mathcal{A}$)
- If $S(\omega) = \sum_{n=1}^{\infty} X_n(\omega)$ for every ω , then S is a r.v. (i.e., $S \in \mathcal{A}$)

The proof of the theorem is quite confusing on Bingyi Jing's book, but Durrett's proof is clear

$sup_i X_i$ is a reconstruction function of countable functions, and logic of the new function is that when we pick out a event ω out of Ω , we calculate a series of function values $X_1(w), X_2(\omega), \ldots$, then we select the maximum value and see it as value of mapping at ω . Then we pick out anther event, over and over again, finally we have new mapping structure from Ω to $\mathcal{R}$ and it's the new function.

If we want to check to see whether it is an r.v. , we use the checking procedure in 3.3

$$\{(sup_i X_i) \leq t\}$$

Since $sup_i X_i$ may not have a closed form, we can't use logic of new function and have to use something simpler. The Maximum value of these function is not greater than t , which means every funtion value is not greater than t , so the above becomes

$$\{sup_i X_i \leq t\} = \cap_{i=1} \{X_i \leq t\}$$

$X_i \in \mathcal{A}$ so $\{X_i \leq t\}$ is an element of σ algebra $\mathcal{A}$. Because σ algebra is closed under countable intersections

$$\{sup_i X_i \leq t\} = \cap_{i=1} \{X_i \leq t\} \in \mathcal{A}$$

so $sup_i X_i$ is an r.v. on $(\Omega, \mathcal{A})$

Apply the same method to $inf_i X_i$

$$\{inf_i X_i \leq t\} = \cup_{i=1} \{X_i \leq t\} \in \mathcal{A}$$

Then we tackle $lim\ sup_i X_i$, $lim\ inf_i X_i$

Using the method we have developed in section 1.3 for $lim\ sup$

$$\lim_{n \to \infty} sup X_n = \lim_{n \to \infty} (sup_{i \geq n} X_i) = inf_{n \geq 1} (sup_{i \geq n} X_i)$$

Since $sup_{i \geq n} X_i$ is proved to be a r.v. , and $inf(something)$ is a r.v. then $lim\ sup_i X_i$ is a r.v.

The same can be applied to $lim\ inf_i X_i$

Then let's prove the second point of the theorem, author's thread is as follow:

$$X(\omega) = \lim_{n} X_n(\omega) = \lim_{n\to\infty} sup X_n(\omega)$$

The reason why the second equal sign is correct is that definition of limit is the convergence of supremum and infimum limit at the same time.

For the third point of theorem

$$S(\omega) = \sum_{n=1}^{\infty} X_n(\omega) = \lim_{n\to\infty} \left(\sum_{i=1}^{n} X_n(\omega) \right)$$

from section 3.4.1, we can see $\sum_{i=1}^{n} X_n(\omega)$ is a r.v., and from the second point of theorem, $\lim_{n\to\infty}(r.v.)_n$ is a .r.v, then $S(\omega)$ is a r.v.

Definition: converges almost surely

Let $X_1, X_2, \ldots$ be a sequence of r.v.'s on $(\Omega, \mathcal{A}, P)$. Define

$$\Omega_0 = \{\omega : lim_n X_n(\omega) exists\} = \{\omega : limsup_n X_n(\omega) - liminf_n X_n(\omega) = 0\}$$

Clearly Ω_0 is measurable. If $P(\Omega_0) = 1$, we say that X_n converges almost surely (a.s.) and write $X_n \to X$ a.s.

First, the space is probability space meaning only on probability space can we mention "converges almost surely". Then we have two r.v.s : $limsup_n X_n$ and $liminf_n X_n$, and last section (Algebraic operations) gives us that subtraction of the two is r.v. .

Since Borel σ algebra can be generated by all finite closed intervals, which means $[0, 0]$ belongs to it, by the definition of r.v., Ω_0 is in $\mathcal{A}$ which means it is measurable.

After we find Ω_0 , if $P(\Omega_0) = 1$, we say that X_n converges almost surely.

The key points of "converge almost surely" : probability space, random variable sequence, limit, measure equals to 1

3.5. Approximations of r.v. by simple r.v.s

In section 3.4, we have dicussed how to construct new variable from existing ones, and there are two ways: finite operations and countably infinite operations.

So how to construct any give function? The answer is this section: give us some bases and infinite operations, we can approximate any function you want.

Theorem:

Given a r.v. $X \geq 0$ on $\Omega, \mathcal{A}$, there exists simple r.v.'s $0 \leq X_1 \leq X_2 \leq \ldots \leq \ldots$ with $X_n(\omega) \uparrow X(\omega)$ for every $\omega \in \Omega$

3.6. σ algebra generated by random variables

3.6.1. Definition

Random variable's corner stone is measurable space, can we do it backwards, meaning using random variable to find a measurable space? The answer is yes.

σ algebra by definition can actually be generated by any class. To put it in plain words, you give me something, I don't care what it is or where it from, but I am going to generate the smallest σ algebra that covers it.

When the source class is somehow related to r.v. , we call the final product σ algebra generated by random variables

Definition:

Let $\{X_\lambda, \lambda \in \Lambda\}$ be a nonempty family of r.v. on $(\Omega, \mathcal{A})$. Define

$$\sigma(X_\lambda, \lambda \in \Lambda) = \sigma(X_\lambda \in B, B \in \mathcal{B}, \lambda \in \Lambda) = \sigma(X_\lambda^{-1}(\mathcal{B}), \lambda \in \Lambda) = \sigma(\cup_{\lambda \in \Lambda} X_\lambda^{-1}(\mathcal{B}))$$

which is called the σ algebra generated by $X_\lambda, \lambda \in \Lambda$

First, for $\Lambda = \{1, 2, \ldots, n\}$ (n may be ∞) , we have

$$\sigma(X_i) = \sigma(X_i^{-1}(\mathcal{B})) = X_i^{-1}(\mathcal{B}) = \{X_i \in \mathcal{B}\}$$
$$\sigma(X_1, \ldots, X_n) = \sigma(\cup_{i=1}^n X_1^{-1}(\mathcal{B})) = \sigma(\cup_{i=1}^n \sigma(X_i))$$

Second, for $\Lambda = \{1, 2, \ldots\}$, it is easy to check that

$$\sigma(X_1) \subset \sigma(X_1, X_2) \subset \ldots \subset \sigma(X_1, \ldots, X_n)$$
$$\sigma(X_1, X_2, \ldots) \supset \sigma(X_2, X_3, \ldots) \supset \ldots \supset \sigma(X_n, X_{n+1}, \ldots)$$

Third, the σ algebra $\cap_{n=1}^\infty \sigma(X_n, X_{n+1}, \ldots)$ is referred to as the tail σ algebra of $X_1, X_2, \ldots$

The logic is each $X_\lambda^{-1}(\mathcal{B})$ can get its own class, and we unionize them to get a new class, and generate a σ algebra based on the new class.

In the first remark, the reason why $\sigma(X_i^{-1}(\mathcal{B})) = X_i^{-1}(B)$ can be established is that in section 3.1, there is a property of mapping: if $\mathcal{B}$ is a σ algebra on Ω_2 , then $X^{-1}(\mathcal{B})$ is a σ algebra in Ω_1

3.6.2. Continuous v.s Discrete

Let's compare checking procedure of random variable and generating σ algebra using random variable, they have one thing in common: using $X^{-1}(\mathcal{B})$.

The former states that since $\mathcal{B}$ is too big, all we have to do is to use X^{-1} on $\mathcal{C}$ that generates $\mathcal{B} = \sigma(\mathcal{C})$. It's the same for the latter: we just can't enumerate all the elements in $\mathcal{B}$ when trying to generate $\sigma(X_i^{-1}(\mathcal{B}))$. Following the same thread as the former, we ask ourselves: can we generate $\sigma(X_i^{-1}(\mathcal{B}))$ just by using some elements in $\mathcal{B}$ and not the whole thing?

$$\sigma(X^{-1}(\mathcal{B})) = \sigma(X^{-1}(\mathcal{C}))\ for\ some\ \mathcal{C} \subset \mathcal{B}$$

It turns out we can! But $\mathcal{C}$ is different between discrete and continuous variables

For discrete random variables X

$$\begin{gathered}
X(\omega) = x_i \ \ 1 \leq i \leq n \\
\mathcal{C} = \{\{x_1\}, \{x_2\}, \ldots, \{x_n\}\} \\
A_i = \{\omega : X(\omega) \in \{x_i\}\} \\
A = \{A_0, A_1, A_2, \ldots, A_n\} \quad (A_0 = \emptyset) \\
\sigma(A) = \sigma(\{A_0, A_1, A_2, \ldots, A_n\})
\end{gathered}$$

By the definition of σ algebra,

$$if\ A_i \in \mathcal{F}\ is\ a\ countable\ sequence\ of\ sets,\ then\ \cup_i A_i \in \mathcal{F}$$

$\sigma(A)$ will contain $\{\cup_{i\in I} A_i : I \subset \{0, 1, \ldots, n\}\}$ meaning

$$\sigma(X^{-1}(\mathcal{C})) = \sigma(A) = \sigma(\{A_0, A_1, A_2, \ldots, A_n\}) = \sigma\left(\{\cup_{i\in I} A_i : I \subset \{0, 1, \ldots, n\}\}\right)$$

Let's look at $\mathcal{B}$, $\mathcal{C} \subset \mathcal{B}$ and $\forall B \in \mathcal{B}$ will cover none or some sets in $\mathcal{B}$, when we use X on $\mathcal{B}$, the result will be

$$\sigma(X^{-1}(\mathcal{B})) = \sigma\left(\{\cup_{i\in I} A_i : I \subset \{0, 1, \ldots, n\}\}\right)$$

so

$$\sigma(X^{-1}(\mathcal{B})) = \sigma(X^{-1}(\mathcal{C}))$$

To put it in words: if we want to generate σ algebra from a discrete r.v., all we have to do is to look at those dicrete values

For continuous random variables X

$\{\omega : X(\omega) = x\}, x \in R$ is not rich enough

$$\mathcal{C} = \{all\ (open, half\ open, closed)\ intervals\}$$

3.7. Distributions and induced distribution functions

In chapter 2, we have talked about measure, meaning based on $(\Omega, \mathcal{C})$, we create a function μ to calculate the value of set in $\mathcal{C}$: $\mu(C), \forall C \in \mathcal{C}$. Specifically in section 2.10, author replaced $\mathcal{C}$ as $\mathcal{F}$, hence $(\Omega, \mathcal{F})$, then restricted it to real number $(\mathcal{R}, \mathcal{B})$. Along with the change of space, we let measure shrink to probability measure and we then have distribution function.

But $(\mathcal{R}, \mathcal{B}, P_X)$ is a special case for probability space, not every problem we study naturally contains real numbers, however they can always be represented by $(\Omega, \mathcal{F}, P)$. How do we go from $(\Omega, \mathcal{F}, P)$ to $(\mathcal{R}, \mathcal{B}, P_X)$?

That is where mapping comes along and the essence of this chapter is to address this issue.

Section 3.1 - 3.6 have addressed $(\Omega, \mathcal{F}) \to (\mathcal{R}, \mathcal{B})$, the part that remains to be solved is $P \to P_X$, and this section handles it

3.7.1. Case I: Random variables

Theorem :

A r.v. X on $(\Omega, \mathcal{F}, P)$ induces another probability space $(\mathcal{R}, \mathcal{B}, P_X)$ through

$$P_X(B) = P(X^{-1}(B)) = P(X \in B),\ \forall B \in \mathcal{B}$$

Definition : X is a r.v.

The distribution of X

$$P_X(B) = P(X^{-1}(B)) = P(X \in B),\ B \in \mathcal{B}$$

The distribution function of X

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

Definition :

- Given two r.v.'s X and Y on $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ respectively, X and Y are identically distributed (i.d.) if $F_X = F_Y$, denoted by $X =_d Y$
- X and Y on $(\Omega, \mathcal{F}, P)$ are equal almost surely (a.s.) if $P(X = Y) = 1$, denoted by $X =_{a.s.} Y$

Definition : A r.v. X on $(\Omega, \mathcal{F}, P)$ is discrete if $\exists$ a countable subset C of $\mathcal{R}$ s.t. $P(X \in C) = 1$

Theorem : X is discrete $\Leftrightarrow$ F_X is discrete

3.7.2. Case II: Random vectors

Definition : $X = (X_1, \ldots, X_n)$ is a random vector

The distribution of X :

$$P_X(B) = P(X^{-1}(B)) = P(X \in B),\ B \in \mathcal{B}^n$$

The (joint) distribution function of X :

$$F_X(x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

3.8. Generating random variables with prescribed distributions

In section 3.7, we have learned the method we use to find P_X from P when we are doing mapping from $(\Omega, \mathcal{F})$ to $(\mathcal{R}, \mathcal{B})$. Essentially the logic is fixing P , X and finding P_X , in other words, P and X controls P_X

Can we do it backwards? using P_X to control the form of X ? The answer is yes

From chapter 2, we know that in probability space $(\mathcal{R}, \mathcal{B}, P_X)$, we can establish a 1-1 correspondence between d.f. F and probability measure, let's rephrase the question: can we use distribution function F to control the form of X ?

Chapter 4 Expectation and Integration

4.1. Definition of Integration

In $(\Omega, \mathcal{F}, \mu)$, we will define $\int f \, d\mu$ for four types of functions: Simple functions, Bounded functions, Nonnegative functions, General functions.

Simple functions: $f(\omega) = \sum_{i=1}^{n} a_i I_{A_i}$

$$\int f \, d\mu = \sum_{i=1}^{n} a_i \mu_i(A_i)$$

Bounded functions: let E be a set with $\mu(E) < \infty$ and let f be a bounded function that vanishes on E^c , To define the integral of f , we observe that if ϕ, ψ are simple functions that have $\phi \leq f \leq \psi$

$$\int f \, d\mu = \underset{\phi \leq f}{sup} \int \phi \, d\mu = \underset{\psi \geq f}{sup} \int \psi \, d\mu$$

Nonnegative functions: if $f \geq 0$, then we let

$$\int f \, d\mu = sup \left\{ \int h \, d\mu : 0 \leq h \leq f, h \; is \; bounded \; and \; \mu(\{x : h(x) > 0\}) < \infty \right\}$$

The logic is stated as follow: there is a nonnagetive function $f \geq 0$, meaning maybe the function goes to ∞ . We want to compute its $\int f \, d\mu$, but don't know how. Since we have defined integration of bounded function, it is natural to pick one out of a series of bounded function integration and use it to represent the intergration of nonnegative function.

1. Based on the nonnegative function f , we can select many functions that satisfy $0 \leq h \leq f$, now h is not necessarily bounded, because f maybe go to infinity.
2. Then we filter all the functions that goes to infinity at some ω .
3. Third, based on what has left at step 2, for each h we find out area $\{x : h(x) > 0\}$ and see if the measure of area is finite. The purpose is to make sure the integration doesn't reach infinity.
4. We have shrink the size of qualified bounded functions, we now use all of them to generate integration values, pick out the biggest one and call it the integration value of f

Though the definition is clear, we still don't know how to calculate the integration of nonnegative function, because the size-shrinked number of qualified bounded functions may still be too large. What should we do? We use limit to approach sub.

Let $E_n \uparrow \Omega$ have $\mu(E_n) < \infty$ and let $a \wedge b = min(a, b)$. Then

$$\int_{E_n} f \wedge n \, d\mu \uparrow \int f \, d\mu \;\; as \; n \uparrow \infty$$

If we want to prove the above lemma, we only have to prove:

$$sup \left\{ \int h \, d\mu \right\} \leq \underset{n \to \infty}{lim \; inf} \int_{E_n} f \wedge n \, d\mu$$

$$\int_{E_n} f \wedge n \, d\mu < \int f \, d\mu$$

The logic is this: $\int_{E_n}$ is integration meaning the result is a number and $lim\ inf_{n\to\infty} = lim_{n\to\infty}(inf_{i\geq n})$ means firstly for every fixed n , we find the infimum of list of numbers after n . We then push n to infinity, the result would be a nondecreasing list, we find the limit of that list. Since $\int_{E_n} f \wedge n\ d\mu$ is a increasing number, the whole thing would essentially becomes

$$\lim_{n\to\infty} \int_{E_n} f \wedge n\ d\mu$$

So the first formula is to prove that no matter which h you choose, the limit of integration of $f \wedge n$ will always bigger than or equal to the integration of h . The second formula shows that the integration $\int_{E_n}$ itself is always smaller that $\int f$, which means when infinity is involved, $\int_{E_n}$ is either smaller than or equal to $\int f$. In other words, $\int_{E_n}$ can only approach $\int f$ from below.

Since the definition is $\int f\ d\mu = sup\{\int h\ d\mu\}$, the only possible explanation is when n is infinity, the three $sup\{\int h\ d\mu\}$, $\int_{E_n} f \wedge n\ d\mu$ and $\int f\ d\mu$ are the same. Hence these two formulas would suffice in order to prove the lemma. So how to prove these fomulas?

The proof of second formula is easy

$$\int_{E_n} f \wedge n\ d\mu = \int (f \wedge n) I_{E_n}\ d\mu = \int h\ d\mu$$

$0 \leq h \leq f$, h is bounded because of n . But I am not sure about $\mu(\{x : h(x) > 0\}) < \infty$ because Durrett only mentioned the measure we use is only σ finite, not finite at the beginning of this section, however let's consider it is true. Then h is a possibility in the sub, which means $\int f$ would be greater than $\int h$ by definition.

For the proof of first formula, let's select a random bounded function h that satisfies the restrictions in the sub, then it must have a upper bound M . Comparing $f \wedge n$ and h we can find that if we want to make sure the former is bigger than or equal to the latter, all we have to do is to choose a $n_0 \geq M$

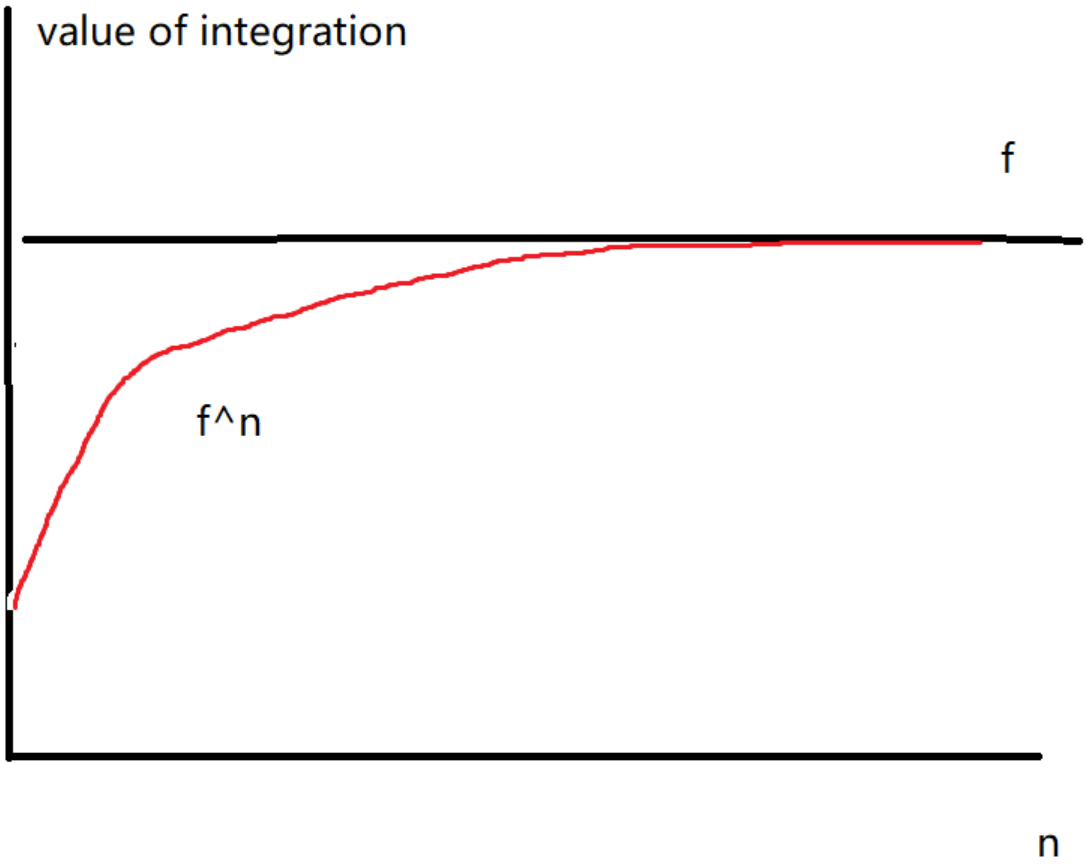
$$n_0 \geq M \Rightarrow f \wedge n_0 \geq h \Rightarrow \int_{E_{n_0}} f \wedge n_0\ d\mu \geq \int_{E_{n_0}} h\ d\mu = \int h\ d\mu - \int_{E_{n_0}^c} h\ d\mu$$

When n increases, E_n^c would shrink, so when we push n to infinity

$$0 \leq \lim_{n\to\infty} \int_{E_n^c} h\ d\mu \leq \lim_{n\to\infty} M\mu(E_n^c \cap \{x : h(x) > 0\}) = 0$$

$$\Rightarrow \lim_{n\to\infty} \int_{E_n^c} h\ d\mu = 0$$

$$\Rightarrow \lim_{n\to\infty}\ inf \int_{E_n} f \wedge n\ d\mu \geq \int h\ d\mu$$



General functions： We say f is integrable if $\int |f| d\mu < \infty$.

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$$

After the definition, we have the following properties for these functions.

$$(1)\ If\ f \geq 0\ a.e.\ then \int f d\mu \geq 0$$

$$(2)\ \forall a \in \mathbf{R}, \int af \, d\mu = a \int f \, d\mu$$

$$(3)\ \int f + g \, d\mu = \int f \, d\mu + \int g \, d\mu$$

$$(4)\ If\ g \leq f\ a.e.\ then \int g \, d\mu \leq \int f \, d\mu$$

$$(5)\ If\ g = f\ a.e.\ then \int g \, d\mu = \int f \, d\mu$$

$$(6)\ \left| \int f \, d\mu \right| \leq \int |f| \, d\mu$$

4.2. Properties of Integration

After defining integration of function, what should we do next? We should study how our action on function effects the property of integration, vice versa. For instance, based on $\int f \, d\mu$, which one is bigger, $\varphi(\int f \, d\mu)$ or $\int \varphi(f) \, d\mu$

There are six theorems in this section: Jensen's inequality, Holder's inequality, Bounded convergence theorem, Fatou's lemma, Monotone convergence theorem, Dominated convergence theorem. The last four all involve limit.

Before introducing Bounded convergence theorem, Durrett presented two types of function convergence in measure theory : convergence in measure, convergence a.e. It's important to see the convergences are based on measure space and they are different from convergences in probability, even though the former can extend to the latter.

$f_n \to f$ *in measure* : for any $\epsilon > 0$

$$\lim_{n\to\infty} \mu(\{x : |f_n(x) - f(x)| > \epsilon\}) = 0$$

$f_n \to f$ *a.e.* :

$$\mu(\{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}) = 0$$

The former is weaker than the latter on a space of finite measure.

Here I want to talk about limit and convergence. We have used limit many time which includes limit of number sequence, limit of set sequence and limit of function sequence. But in this chapter, we encounter heavy usage of function convergence. In a nutshell, limit is just a tool we use to check convergence.

For a function, if

$$\lim_{n\to\infty} sup f_n = \lim_{n\to\infty} inf f f_n$$

it actually implies that every point x forms a number sequnce, and the above formula holds true for every number sequencewe. We call it $lim_{n\to\infty} f_n = f$. So the above is a very restricted situation because sup=inf has to be true for every x .

However when we talk about convergence in measure theory, things are slightly loose than limit of funtion. Take a.e. for example: $\mu(\{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}) = 0$. There are two types of bad points:

1. Some points don't have limit at all
2. Even though all points have limit, but their values don't equal to their values on f

However as long as the measure value of those bad points is zero, it's safe to say f_n converge to f , or $f_n \to f$

So function limit doesn't guarantee function convergence, it depends on the form of f ; and function convergence doesn't guarantee function limit due to those measure-zero bad points; limit is just a tool we use to check convergence.

Bounded convergence theorem : Let E be a set with $\mu(E) < \infty$. Suppose f_n vanishes on E^c , $|f_n(x)| \leq M$, and $f_n \to f$ in measure. Then

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu$$

Fatou's Lemma : If $f_n \geq 0$ then

$$\underset{n\to\infty}{lim\ inf} \int f_n \, d\mu \geq \int (\underset{n\to\infty}{lim\ inf}\ f_n) \, d\mu$$

Monotone convergence theorem : If $f_n \geq 0$ and $f_n \uparrow f$ then

$$\int f_n \, d\mu \uparrow \int f \, d\mu$$

Dominated convergence theorem : If $f_n \to f \ a.e.$, $|f_n| \leq g$ for all n , and g is integrable, then

$$\int f_n \, d\mu \to \int f \, d\mu$$

4.3. Expected Value

$(\Omega, \mathcal{F}, P)$, P is a probability measure and X is a random variable.

4.4. Moment Generating Function and Characteristic Function

Mathematical forms of the two are roughly the same: $E[e^{tx}]$ and $E[e^{itx}]$

- Moment Generating Function can determine distribution (i.e. it contains all the information of a distribution) and can help us to calculate moments
- Characteristic Function can also determine distribution and can help to prove CLT.

The main usage of MGF is to derive distribution and the main usage of CF is to prove CLT. It is why Durrett introduces CF in CLT chapter.

Chapter 5 Laws of Large Numbers

5.1. Independence

5.1.1. Definition

Independence of event :

In $(\Omega, \mathcal{F}, P)$, there are many events that belong to $\mathcal{F}$.

- If we take two specific A, B out of $\mathcal{F}$ and find out $P(A \cap B) = P(A)P(B)$, we can say A, B are independent.
- If we take n events out of $\mathcal{F}$ and find out no matter how we do permutation and combination , we always have $P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i), I \subset \{1, \dots, n\}$, we say these events are independent.
- Infinite collection of events is said to be independent if every finite subcollection is.

One thing is important : the expansion of the first point is called pair-wise independence. multi-events independence is pair-wise independence and the converse is not true.

Independence of random variable :

In $(\Omega, \mathcal{F}, P)$, we can put many functions on $\mathcal{F}$

- Two function X, Y , we always have $P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$, no matter how we select $C, D \in \mathcal{R}$
- n functions, we always have $P(\cap_{i=1}^{n} \{X_i \in B_i\}) = \prod_{i=1}^{n} P(X_i \in B_i)$, no matter how we select $B_i \in \mathcal{R}$
- Infinite collection of events is said to be independent if every finite subcollection is

The restriction of independence for r.v.s is more than that for events. Like I said before, the reverse function has shrinking effect, so when we do $X \in C$, the resulting sets would be a part of $\mathcal{F}$. What author says here is that as long as the measure of these sets satisfies the above restriction, we can call these r.v.s independent.

Independence of σ field :

There are more than one measure space : $(\Omega, \mathcal{F}_i, P)$, which means based on one Ω and one P , we have more than one σ field

- Two σ fields $\mathcal{F}, \mathcal{G}$, we always $P(A \cap B) = P(A)P(B)$ no matter how we select $A \in \mathcal{F}, B \in \mathcal{G}$

- n σ fields $\mathcal{F}_i$, we always have $P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i)$ no matter how we select $A_i \in \mathcal{F}_i$
- Infinite collection of events is said to be independent if every finite subcollection is

The restriction here is more than that of last one. Since reverse function has shrinking effect, only a portion of sets of $\mathcal{F}_i$ would do the trick. But here we have multiple σ fields, for us to say they are independent, the measure of every set has to satisfy the above restriction, not just some sets.

5.1.2. Two types of production

Without doubt, our main focus must be r.v.s , but there are shit loads of sets in Borel sets, if we check independence by definition, it'll never ends. Recall how we check a random variable, we didn't go through every set, we only check the thing of which $\mathcal{R}$ is the smallest σ algebra. Here we will do the same thing.

But first we need to generalize our previous three definitions, because it's just exhausting to do all the steps three times every time we want do something about independence.

Collections of sets $\mathcal{A}_1, \mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{F}$ are said to be independent if whenever $A_i \in \mathcal{A}_i$ and $I \subset \{1, \ldots, n\}$ we have

$$P(\cap_{i\in I} A_i) = \prod_{i\in I} P(A_i)$$

Why this can be general form for all three? Let's check one by one.

For independence of event, we can put only one event in every $\mathcal{A}_i$, then it's exactly the same as the original definition.

For random variable and σ field, we notice here we use $\prod_{i\in I}$ but before we use $\prod_{i=1}^n$. If we have n collections of sets. The latter picks one set from every collection we have listed, and the former not only does what the latter does, but also pick one set from every collection of some collections that have been listed. So the number of actions that is suggested by the former is larger than the latter. It seems like the the definition here can not be generalized to random variable and σ field, but actually it can. The reason is when we discuss random variable, the collection of sets $\mathcal{A}$ we find through funtion usually contains Ω , and a σ field naturally contains Ω

$$P(\cap_{i\in I} A_i) = P(\cap_{i\in I} A_i \cap_{j\in I^c} \Omega) = \prod_{i\in I} P(A_i) \prod_{j\in I^c} P(\Omega) = \prod_{i\in I} P(A_i)$$

So when I does not contain all the integers from 1 to n , but only a part of them, what $\prod_{i\in I}$ does is to random select set from the collections that are mentioned by I , and only select Ω from the collections that are not mentioned by I . But this situation is already coverd by the I that contains all the integers from 1 from n , which is also $\prod_{i=1}^{n}$. So as long as we have Ω in every collection, even though $\prod_{i\in I}$ is redundant compared with $\prod_{i=1}^{n}$, they are essentially equal to each other.

When $\mathcal{A}_i$ is found through function, the above definition is the definition of independence for random variable. When $\mathcal{A}_i$ is σ field, the above definition is the definition of independence for σ field (the fact all the $\mathcal{A}_i$ is the subset of a giant σ field is a little weird)

5.1.3. Checking procedure

After the arduous journey discussing generalization and a bunch of other crap, we derive the following theorem

Theorem : In order for $X_1, \ldots, X_n$ to be independent, it is sufficient that for all $x_1, \ldots, x_n \in (-\infty, \infty]$

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i)$$

It means we don't have to check every subset of R , only the subset with the form $(-\infty, x_i]$. It is exactly how we did when checking a random variable. However we are not satisfied, becuase the above only states how we should check independence by distribution function. How about density functions and discrete random variables?

Corollary 1 : Suppose $(X_1, \ldots, X_n)$ has density $f(x_1, x_2, \ldots, x_n)$, that is

$$P((X_1, X_2, \ldots, X_n) \in A) = \int_A f(x)dx \ for \ A \in \mathcal{R}^n$$

If $f(x)$ can be written as $g_1(x) \ldots g_n(x)$ where the $g_m \geq 0$ are measurable, then $X_1, X_2, \ldots, X_n$ are independent (Note that the g_m are not assumed to be probability densities)

Corollary 2 : Suppose $X_1, \ldots, X_n$ are random variables that take values in countable sets $S_1, \ldots, S_n$. Then in order for $X_1, \ldots, X_n$ to be independent, it is sufficient that whenever $x_i \in S_i$,

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

Having stated how individual random variables are independent from each other, we consider the independence of their combination.

Theorem : If for $1 \leq i \leq n, 1 \leq j \leq m(i)$, $X_{i,j}$ are independent and $f_i : \mathbf{R}^{\mathbf{m(i)}} \to \mathbf{R}$ are measurable, then $f_i(X_{i,1}, \ldots, X_{i,m(i)})$ are independent.

5.1.4. Distribution and Expectation

Theorem : Suppose $X_1, \ldots, X_n$ are independent random variables and X_i has distribution μ_i . Then $(X_1, \ldots, X_n)$ has distribution $\mu_1 \times \ldots \times \mu_n$.

Here I want to clarify why measure would change with random variables. When discussing multiple random variables, we have a fixed stage $(\Omega, \mathcal{F}, P)$, meaning if you want to find the measure value of certain set on $\mathcal{F}$, you have to use P . But μ_i doesn't focus on $\mathcal{F}$, it focuses on $\mathbf{R}$, that's why it changes with X_i

$$\mu_i(r) = P(X_i^{-1}(r)) \; for \; r \in \; \mathbf{R}$$

So in the future, when we mention the distribution of X , it means $P(X^{-1}(r))$

Theorem : Suppose X and Y are independent and have distributions μ and ν . if $h : \mathbf{R}^2 \to \mathbf{R}$ is a measurable function with $h \geq 0$ or $E\,|h(X, Y)| < \infty$, then

$$Eh(X, Y) = \int \int h(x, y) \mu(dx) \nu(dy)$$

5.2. Weak Laws of Large Numbers

Definition : We say that X_n converges to X in probability if for all $\epsilon > 0$, $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$

5.2.1. L^2 Weak Laws

Chapter 6 Central Limit Theorem

6.1. Connection between LLN and CLT

Laws of Large Number indicates that it is about a number, Central Limit Theorem is about a distribution. They both discuss what would happen of $\overline{X}$ when $n \to \infty$.

Distribution of $\overline{X}$ would approach normal distribution (CLT) , and as n gets even bigger the variance of such distribution would decrease and $\overline{X}$ approaches a fixed value (LLN) .

- LLN is a rougher version of CLT (without information of variance and it is justified because here n is really big even for CLT)
- CLT is a finer version of LLN (with information of variance)