Note 3_1: measure theory & stochastic differential equation

Yining He

03/31/20

1) Terminology

1.1 "Mathematically ---- Divergence": Wiki (https://en.wikipedia.org/wiki/Divergence)

Definition in coordinates [edit]

Cartesian coordinates [edit]

In three-dimensional Cartesian coordinates, the divergence of a continuously differentiable vector field $\mathbf{F} = F_x\mathbf{i} + F_y\mathbf{j} + F_z\mathbf{k}$ is defined as the scalar-valued function:

$$\operatorname{div}\mathbf{F} = \nabla\cdot\mathbf{F} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\cdot(F_x, F_y, F_z) = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.$$

Although expressed in terms of coordinates, the result is invariant under rotations, as the physical interpretation suggests. This is because the trace of the Jacobian matrix of an N-dimensional vector field $\mathbf{F}$ in N-dimensional space is invariant under any invertible linear transformation.

1.2 "Markov model": Wiki (https://en.wikipedia.org/wiki/Markov_model)

"*In probability theory, a Markov model is a stochastic model used to model randomly changing systems. It is assumed that future states depend only on the current state, not on the events that occurred before it (that is, it assumes the Markov property). Generally, this assumption enables reasoning and computation with the model that would otherwise be intractable*."

Notes: "Markov chain": Wiki (https://en.wikipedia.org/wiki/Markov_chain)

"***A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event***. *In continuous-time, it is known as a Markov process. Markov chains have many applications as statistical models of real-world processes, such as studying cruise control systems in motor vehicles, queues or lines of customers arriving at an airport, currency exchange rates and animal population dynamics. Markov processes are the basis for general stochastic simulation methods known as Markov chain Monte Carlo, which are used for simulating sampling from complex probability distributions, and have found application in Bayesian statistics and artificial intelligence. The adjective Markovian is used to describe something that is related to a Markov process.*"

04/01/20

1) Terminology

1.1 "齐次微分方程 or Homogeneous differential equation"

(A) Wiki https://en.wikipedia.org/wiki/Homogeneous_differential_equation

A differential equation can be **homogeneous** in either of two respects.

A first order differential equation is said to be homogeneous if it may be written

$$f(x,y)dy = g(x,y)dx,$$

where f and g are homogeneous functions of the same degree of x and y.[1] In this case, the change of variable $y = ux$ leads to an equation of the form

$$\frac{dx}{x} = h(u)du,$$

which is easy to solve by integration of the two members.

Otherwise, a differential equation is homogeneous if it is a homogeneous function of the unknown function and its derivatives. In the case of linear differential equations, this means that there are no constant terms. The solutions of any linear ordinary differential equation of any order may be deduced by integration from the solution of the homogeneous equation obtained by removing the constant term.

Homogeneous linear differential equations [edit]

See also: Linear differential equation

A linear differential equation is **homogeneous** if it is a homogeneous linear equation in the unknown function and its derivatives. It follows that, if $\phi(x)$ is a solution, so is $c\phi(x)$, for any (non-zero) constant c. In order for this condition to hold, each nonzero term of the linear differential equation must depend on the unknown function or any derivative of it. A linear differential equation that fails this condition is called **inhomogeneous.**

A linear differential equation can be represented as a linear operator acting on *y(x)* where *x* is usually the independent variable and *y* is the dependent variable. Therefore, the general form of a linear homogeneous differential equation is

$$L(y) = 0$$

where *L* is a differential operator, a sum of derivatives (defining the "0th derivative" as the original, non-differentiated function), each multiplied by a function f_i of *x*:

$$L = \sum_{i=0}^{n} f_i(x) \frac{d^i}{dx^i},$$

where f_i may be constants, but not all f_i may be zero.
For example, the following linear differential equation is homogeneous:

$$\sin(x)\frac{d^2y}{dx^2} + 4\frac{dy}{dx} + y = 0,$$

whereas the following two are inhomogeneous:

$$2x^2\frac{d^2y}{dx^2} + 4x\frac{dy}{dx} + y = \cos(x);$$

$$2x^2\frac{d^2y}{dx^2} - 3x\frac{dy}{dx} + y = 2.$$

The existence of a constant term is a sufficient condition for an equation to be inhomogeneous, as in the above example.

(B) “齐次方程的「齐次」代表什么？” https://www.zhihu.com/question/23599338

1.首先看 $\frac{dy}{dx}=\phi(\frac{y}{x})$

把这个方程变形一下，变成 $\frac{dy}{dx}-\phi(\frac{y}{x})=0$

然后我们可以利用等号左边得到函数 $f(x,y)=\frac{dy}{dx}-\phi(\frac{y}{x})$

然后就会发现，f(x,y)是符合最前面齐次函数的定义的。

$$f(ax,ay)=\frac{d(ay)}{d(ax)}-\phi(\frac{ay}{ax})=a^0(\frac{dy}{dx}-\phi(\frac{y}{x}))=\frac{dy}{dx}-\phi(\frac{y}{x})$$

这是一个0次齐次的情况。

其实本质上，"齐次"这个词，"次"代表着函数中每一个独立"项"的指数，"齐"表示一样。函数是"齐""次"的，现在翻译出来，即"函数的每一个独立项的指数是相等的"，这与函数满足 $f(ax)=a^k f(x)$ （齐次的定义），互为充要条件。

比如 $f(x,y)=xy+x^2$ ，也是一个齐次的（函数）。

而 $f(x,y)=xy+x^2+5$ ，就不是齐次，因为常数等于 $5x^0$ ，这一项的指数不同于其他项，所以它就不是齐次，就没有办法得到 $f(ax)=a^k f(x)$ 这个规律。

关于项的指数，要看所有的指数和。如xy，对于x和y都是一次，但是二者相乘，指数相加，为2。所以 xy,x^2,y^2 是同指数，也就是同次的，是"齐""次"的，是齐次的。

综上， $\frac{dy}{dx}=\phi(\frac{y}{x})$ 被称为齐次方程，能够发现，它拥有"其相关二元函数"

$f(ax,ay)=\frac{d(ay)}{d(ax)}-\phi(\frac{ay}{ax})=a^0(\frac{dy}{dx}-\phi(\frac{y}{x}))=\frac{dy}{dx}-\phi(\frac{y}{x})$ 是齐次函数的特性。

2.再看 $\frac{dy}{dx}+P(x)y=Q(x)$ **当** $Q(x)$ **恒等于0时为齐次**

我们仍旧可以得到它的相关二元函数

$$g(x,y)=\frac{dy}{dx}+P(x)y-Q(x)$$

然后发现，如果按照上一部分的说法，

这里 $\frac{dy}{dx}$ 就成了0次，而P(x)y是2次，即使Q(x)=0,仍然是有两项不同次，是不齐次的。为什么也被叫齐次呢?

这确实就是不同的地方。

因为这里，对于n阶线性微分方程，齐次的考量不再是g(x,y)，而变成了g(y) ——至于为什么变量数量考虑上不同，在后文"额外"部分有写一些思考。

这时候就会发现，g(y)是具有齐次性的

当Q(x)恒等于0时：

$$g(ay)=\frac{d(ay)}{dx}+P(x)ay-0=a[\frac{dy}{dx}+P(x)y]=ag(y)$$ ，齐次

当Q(x)不等于0时：

齐次条件不成立，非齐次

综上所述，应该能够感受到这其中的"齐次"本质有相似的地方，但是考量的对象确实是不同的。有的时候用的是对应的f(x,y)，有的时候只用g(y)。

3.额外：关于二者"不同"的思考

我是从实用角度，从"这样设定的目的"出发来理解的。

"齐次方程"一节，一开始说的是

一、齐次方程

如果一阶微分方程可化成

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\varphi\left(\frac{y}{x}\right) \tag{1}$$

的形式，那么就称这方程为齐次方程，例如

$$(xy-y^2)\mathrm{d}x-(x^2-2xy)\mathrm{d}y=0$$

是齐次方程，因为它可化成

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{xy-y^2}{x^2-2xy},$$

即

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\frac{y}{x}-\left(\frac{y}{x}\right)^2}{1-2\left(\frac{y}{x}\right)}.$$

这一章介绍的齐次方程，不局限于线性微分方程，还包括非线性的微分方程，比如这样的：

例 1 解方程

$$y^2+x^2\frac{\mathrm{d}y}{\mathrm{d}x}=xy\frac{\mathrm{d}y}{\mathrm{d}x}.$$

解 原方程可写成

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{y^2}{xy-x^2}=\frac{\left(\frac{y}{x}\right)^2}{\frac{y}{x}-1},$$

而这一章的目的，是讲述一个通用的方法，能够线性非线性都求的。那么这个方法，就是要依靠 $\frac{y}{x}=u$ 这个形式的替换。这样，这个方法涉及到了y,x**两个量**，且**二者**之间有紧密的运算关系，是一个**整体**，所以就考虑**这二者**相互联系的齐次性，所以**f(x,y)，二元**函数。

而对于“一阶线性微分方程”一章，并不着眼于一个同时拥有y,x的项，这一章，限定在“线性”的条件下。

然后现在我们再来看这里的“一阶线性微分方程”一节

首先我们要理解一下，这里方程的“线性”和前面说的那个函数的线性是什么关系

这其实很关键。

因为实际上在讨论一个微分方程是不是线性的时候，

一、线性方程

方程

$$\frac{\mathrm{d}y}{\mathrm{d}x}+P(x)y=Q(x) \qquad (1)$$

叫做一阶线性微分方程，因为它对于未知函数 y 及其导数是一次方程．如果 $Q(x)\equiv0$，则方程(1)称为齐次的；如果 $Q(x)\not\equiv0$，则方程(1)称为非齐次的．

数学工作者并不把这个方程的右边清零，而是把没有y的项放到右边。是不是线性，是以等号左边的代数式作为的函数 $h(y)=\frac{dy}{dx}+P(x)y$ 来理解的。且只着眼于y。

可以通过检验知道h(y)是线性的。

这也就是“齐次”的用法出现分歧的地方。

因为在考虑线性的时候，就使用了这个h(y)，而不是h(x,y)。齐次作为之后考虑的一个性质，也遵从了这一点。

但是！齐次有一点是统一的，即考虑齐次与否的时候还是对方程先整理成等于0的形式：

$$\frac{dy}{dx}+P(x)y-Q(x)=0$$

然后讨论函数 $g(y)=\frac{dy}{dx}+P(x)y-Q(x)$ 是否齐次

显然Q(x)恒等于0时，符合齐次（这里是一次齐次，注意!这里只讨论y).

我天，我认为这才是对高等数学微分方程齐次和线性部分这两节内容的一个能说通的理解。

内容精炼：

"齐次"概念的源头是齐次函数： $f(ax)=a^k f(x)$

要理解方程的齐次，就要从方程中"找到"这样的函数！

试着将 $\frac{dy}{dx}=\phi\frac{dy}{dx}$ 转化为 $f(x,y)=\frac{dy}{dx}-\phi\frac{dy}{dx}$ 看看是不是齐次（函数）？

试着将 $\frac{dy}{dx}+P(x)y=Q(x)$ 转化为 $g(y)=\frac{dy}{dx}+P(x)y-Q(x)$ 看看是不是齐次（函数）？
（请结合高数课本，有一些更具体的条件比如Q(x)是否恒等于0等）

如果奇怪为什么有的时候转化为f(x,y)，有的时候转化为f(y)?

再去看我关于方程的线性和函数的线性的解释吧！

Notes: 常微分方程(https://en.wikipedia.org/wiki/Ordinary_differential_equation)

在数学分析中，**常微分方程**（英语：ordinary differential equation，简称ODE）是未知函数只含有一个自变量的微分方程。对于微积分的基本概念，请参见微积分、微分学、积分学等条目。

很多科学问题都可以表示为常微分方程，例如根据牛顿第二运动定律，物体在力的作用下的位移 s 和时间 t 的关系就可以表示为如下常微分方程：

$$m\frac{d^2s}{dt^2}=f(s);$$

其中 m 是物体的质量，$f(s)$ 是物体所受的力，是位移的函数。所要求解的未知函数是位移 s，它只以时间 t 为自变量。

04/05/20

1) "σ-代数" (σ-algebra): Wiki (中文)

在数学中，某个集合 X 上的 **σ–代数**又叫 **σ–域**，是 X 的幂集的子集合（X 的幂集即包含所有 X 的子集的集合系）。这个子集满足对于补集运算和可数个并集运算的封闭性（因此对于可数个交集运算也是封闭的）。σ–代数在测度论里可以用来定义所谓的"**可测集合**"，是测度论的基础概念之一。

定义 [编辑]

让 X 为非空集合，集合系 $\mathcal{F}$ 中的元素是 X 的子集合，满足以下条件的集合系 $\mathcal{F}$ 称为 X 上的一个 **σ–代数**：[1][2]

- X 是集合系 $\mathcal{F}$ 中的元素；
- 如果集合 A 在 $\mathcal{F}$ 中，那么它的补集 A^c 也在 $\mathcal{F}$ 中；
- 如果有可数个集合 $A_1, A_2, \cdots$ 都在 $\mathcal{F}$ 中，那么它们的并集也在 $\mathcal{F}$ 中。

以上条件用数学语言来表示，就是：

X 为一集合，假设有集合系 $\mathcal{F} \subseteq \mathcal{P}(X)$，其中 $\mathcal{P}(X)$ 代表 X 的幂集，若 $\mathcal{F}$ 满足下列条件

- $X \in \mathcal{F}$;
- $A \in \mathcal{F} \;\Rightarrow\; A^c \in \mathcal{F}$;
- $A_n \in \mathcal{F},\ \forall n \in \mathbb{N} \;\Rightarrow\; \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

则称集合系 $\mathcal{F}$ 是 X 的 σ–代数。

在测度论里 $(X, \mathcal{F})$ 称为一个可测空间。集合族 $\mathcal{F}$ 中的元素，也就是 X 的某子集，称为可测集合。而在概率论中，这些集合被称为随机事件。

例子 [编辑]

- 有两个σ–代数的简单例子，它们分别是：
  1. X上含集合最少的σ–代数$\{\emptyset, X\}$；和
  2. X上含集合最多的σ–代数是X的幂集$2^X := \{A : A \subset X\}$。
- 假设集合$X = \{a, b, c, d\}$，那么$\mathcal{F} = \{\varnothing, \{a\}, \{b, c, d\}, X\}$ 是集合X上的一个σ–代数。这也是所有包含$\{a\}$的σ–代数中最“小”的一个。

性质 [编辑]

σ–代数是一个代数（域）也是一个λ系，它对集合的交集、并集、差集、可数交集、可数并集运算都是封闭的。

2) “Measure (mathematics)” （测度）

(I) Wiki (中文)

数学上，**测度**（英语：**measure**）是一个函数，它对一个给定集合的某些子集指定一个数，这个数可以比作大小、体积、概率等等。传统的积分是在区间上进行的，后来人们希望把积分推广到任意的集合上，就发展出测度的概念，它在数学分析和概率论有重要的地位。

测度论是实分析的一个分支，研究对象有σ代数、测度、可测函数和积分，其重要性在概率论和统计学中都有所体现。

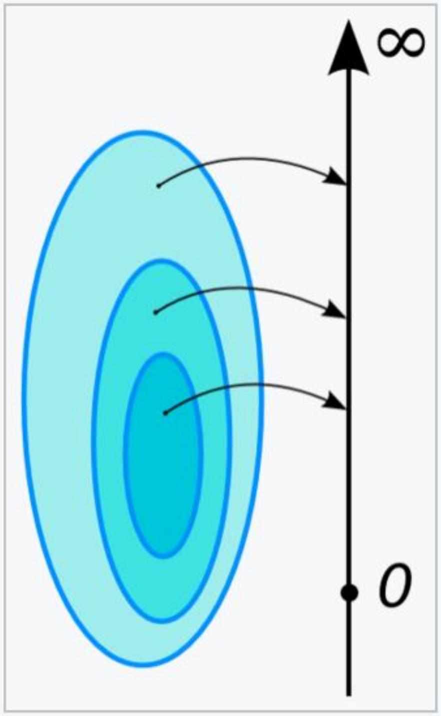
定义 [编辑]

X是个集合，定义在 X上的另一集合 $\mathcal{A}$ ，$\mathcal{A}$中的元素是 X的子集合，而且是一个σ–代数，测度 μ（详细的说法是**可数可加的正测度**）是个定义在 $\mathcal{A}$ 上的函数，于$[0,\infty]$中取值，且满足以下性质：

- **非负性质**：对所有的 $E \in \mathcal{A}$，有 $\mu(E) \geq 0$，
- **空集合的测度为零**：$\mu(\varnothing) = 0$，
- **可数可加性**，或称 **σ–可加性**：若 $\{E_k\}_{k=1}^{\infty}$ 为 $\mathcal{A}$ 中可数个两两不相交元素的集合，换句话讲，对所有 $E_i, E_j \in \{E_k\}_{k=1}^{\infty}$，$i \neq j$ 有 $E_i \cap E_j = \varnothing$，则可得

  $$\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)。$$

这样的三元组$(X, \mathcal{A}, \mu)$称为一个**测度空间**，而$\mathcal{A}$ 中的元素称为这个空间中的**可测集合**。



通俗的说，测度把每个集合映射到非负实数来规定这个集合的大小：空集的测度是0；集合变大时测度至少不会减小（因为要加上变大的部分的测度，而它是非负的）。

σ-有限测度 [编辑]

主条目：σ–有限测度

如果$\mu(X)$ 是一个有限实数（而不是∞），则测度空间$(X, \mathcal{A}, \mu)$称为**有限测度空间**。非零的有限测度与概率测度类似，因为可以通过乘上比例因子$\frac{1}{\mu(X)}$进行归一化。如果X 可以表示为可数个可测集的并集，而且这些可测集的测度均有限，则该测度空间称为**σ–有限测度空间**。如果测度空间中的一个集合A 可以表示为可数个可测集的并集，而且这些可测集的测度均有限，就称A **具有σ–有限测度**。

作为例子，实数集赋以标准勒贝格测度是σ–有限的，但不是有限的。为说明之，只要考虑闭区间族[k, k+1]，k取遍所有的整数；这样的区间共有可数多个，每一个的测度为1，而且并起来就是整个实数集。作为另一个例子，取实数集上的计数测度，即对实数集的每个有限子集，都把元素个数作为它的测度，至于无限子集的测度则令为∞。这样的测度空间就不是σ–有限的，因为任何有限测度集只含有有限个点，从而，覆盖整个实数轴需要不可数个有限测度集。σ–有限的测度空间有些很好的性质；从这点上说，σ–有限性可以类比于拓扑空间的可分性。

(II) (概率论和实变函数（测度论）有什么联系?) https://www.zhihu.com/question/29800166

(a)

好吧还是展开说一说。按照我本科和PhD所在学校的教学设置，在没有测度论的前提下，一般可以开概率论和应用随机过程。这些课会包含古典/几何概型，常见分布，不证明的大数定律和中心极限定律，马氏链，泊松过程，条件期望和鞅，甚至一点点布朗运动。对不以随机分析和花式scaling limit为方向的人来说，这些已经足够开始科研了。但其实这里很多事情我们都说不清：比如连续变量的条件概率，比如马氏过程常返性中涉及的无穷样本轨道，比如强大数定律a.s.和i.o.......而测度论算是填上了这个背景里的坑。

又或者布朗运动是定义在全体连续函数上的Wiener测度，但几乎处处考虑的都是处处不可微的连续函数，我不知道有多少分析的人，会对这样性质不友好的函数感兴趣。概率里会有不变原理，会有重对数律。

(b)

测度论忘得差不多了……就还记得测度空间的那三个元素……

举几个简单的例子：

1.扔一个fair coin（公平的硬币？），出现正面和出现反面的概率都是二分之一。基本的概率论就能解释，很直观的“概率论”。

2.我要是扔无穷次硬币呢？在扔无穷次硬币的情况下，实际上可能出现的情况也是无穷多种的，那么特定的情况发生的概率（比如出现全是正面） $P=\frac{1}{\infty}=0$.

3.还是扔无穷次硬币的情况，我想要前三次都是正面的概率要怎么算？因为每一个独立的情况出现的概率都是0，是不是我就要把这些0加一起呢？结果还是0？很明显不是。

在这种情况下，就需要用上测度论了。在我看来测度论就是你有一个样本空间，里面是所有可能发生的随机事件，然后你有一个函数P（概率测度）将样本空间里的某些子集映射到实轴上某条连续的线段上（准确的说应该叫borel set，博雷尔子集，一般是[0,1]内的某个点或者某条线段吧）。这样做的最大好处是可以将离散的事件对应到“连续的”一个集合（线段）当中，有了连续性就可以随便操作了，加减乘除随意。而且每一个实轴上的线段（子集）都有与之对应的样本空间的意义。

(c)
(c)-1. (概率是什么？Sigma algebra, Borel field 是什么意思，意义何在？)
https://www.zhihu.com/question/37316236

张敬信
高校数学教师，不解答中学数学问题，不解答智力题

我从概率论和测度论的关系的角度给你解释一下啊，测度论是概率论的理论基础，所以概率中的一些概念抽象化就是对应的测度论中的概念。

概率是要度量“事件发生的可能性”的大小，事件的抽象化描述就是集合，需要考察“事件的全体”，对应到测度论就是“集合系”。“事件发生的可能性“是对事件的一种度量，对应到测度论就是“集合的测度”。

不是每个事件都可以定义其概率（发生的可能性的大小）的，对应的就是不是每个集合都可以定义测度，可以定义测度集合就是可测集。同时，事件必然要涉及到事件的组合运算（复杂事件是可由基本事件表示出来），对应的就是集合的交、并、差、余、极限的运算到复杂集合，所以又需要保证做可列次这些运算不能超出全体范围（即可测集的范围要足够大，以保证集合的可列次交、并、差、余、极限的运算，之后还在里面）

那么什么样的集合系，才能保证其中的集合是可测集（可以定义测度，又对那些运算封闭）呢？测度论中讲了，只要集合系是σ-代数（也叫σ-域）就可以了。σ-代数的基本定义是：1. 全集在里面；2. 里面每个集合的余集在里面；3. 里面任意可列个集合的并集在里面。有了这三条基本定义，就可以推出：空集、可列次交、并、差、上限集、下限集运算之后都能在里面。就满足需要了。

所以，集合X+该集合上的一个σ-代数F，（X，F）就是一个可测空间了，即可以定义测度的空间（F中任一集合都可以定义其测度（某种度量））。进一步再定义了测度μ，那么（X, F, μ）就是测度空间。

对应到概率论中，样本空间Ω，事件域F（是个σ-代数），概率测度P，放一起（Ω，F，P）就是概率测度空间。概率测度P是满足特殊要求的一种测度：$P(\Omega)=1$.

Borel Feild就是Borel σ-代数，表示实数轴上的σ-代数，可由实轴上的所有开集生成（的σ-代数），也可由实数轴上所有的$(-\infty,a]$这样的区间生成（σ-代数），是相等的。按σ-代数前面说的，实数轴上开集、闭集的至多可列次交、并、差(余)、上限集、下限集、极限集的运算，都超不出该Borel σ-代数的范围。

Borel σ-代数（我用Br表示）有什么用？其实概率论中的随机变量、对应测度论中的可测函数，而可测函数就是从可测空间（X，F)到（R，Br）的可测映射。

再说说随机变量，前面说了概率论中要用集合表示事件，但事件五花八门，怎么统一用一种简单的集合表示呢？这就用到映射的概念，建立一种从样本空间（基本事件的全体）到实数轴的映射（一一对应）就可以了，这种映射就是随机变量。有了它，基本事件映射到实数轴上就是的基本区间，基本事件经过运算生成的复杂事件，映射到实数轴上就是实数轴上Borel σ-代数中的集合。

因为有了这个对应关系，要度量“事件发生的可能性的大小”（即概率测度），只要度量“实数轴上Borel σ-代数中的集合” 就可以了（前面说了Br因为是σ-代数是可以定义测度的，给Br中的集合定义概率测度就行了）。

所以，随机变量的测度论语言定义是这样的：设（Ω，F，P）为概率测度空间，若对实数轴上Borel σ-代数中的任一集合（称为Borel集）B，都有 {w∈Ω: X(w) ∈B} ∈F，则称X(w)为随机变量。

总之，随机变量就是建立了“随机事件”到“实数轴上Borel σ-代数”的一种对应，并且保证了建立了这种对应的随机事件都是可以定义概率测度的。

既然随机事件{w∈Ω: x(w) ∈B}属于F，那么可以有概率，即P{w∈Ω: x(w) ∈B}是有意义的，为了简单，概率中就记P{w∈Ω: X(w)∈B} = P{X ∈B} 了。

特别地，若取B=(-∞,x), 则事件{X∈B}的概率

P{X∈B} = P{X≤x} := F(x)

就定义成随机变量X的分布函数。因为对任意的区间(a,b], 都可表示成

P{X ∈(a,b] } = P{a<X≤b} = P{X≤b} - P{X≤a} = F(b)-F(a)

进而，由这样的区间经过至多可列次交、并、差运算的复杂的实数轴上的Borel集都可以用F(x)给出其概率。

当然，随机变量也可以定义为从样本空间到平面R2上的映射，就是二元随机变量。

(c)-2.

我回答另一个知乎问题【概率是什么？Sigma algebra, Borel field 是什么意思，意义何在？】刚好也适合这个问题，再补充一部分给你：

测度论是概率论的理论基础，所以概率中的一些概念抽象化就是对应的测度论中的概念。

概率是要度量“事件发生的可能性”的大小，事件的抽象化描述就是集合，需要考察“事件的全体”，对应到测度论就是“集合系”。”事件发生的可能性“是对事件的一种度量，对应到测度论就是“集合的测度”。

不是每个事件都可以定义其概率（发生的可能性的大小）的，对应的就是不是每个集合都可以定义测度，可以定义测度集合就是可测集。同时，事件必然要涉及到事件的组合运算（复杂事件是可由基本事件表示出来），对应的就是集合的交、并、差、余、极限的运算到复杂集合，所以又需要保证做可列次这些运算不能超出全体范围（即可测集的范围要足够大，以保证集合的可列次交、并、差、余、极限的运算，之后还在里面）
那么什么样的集合系，才能保证其中的集合是可测集（可以定义测度，又对那些运算封闭）呢？测度论中讲了，只要集合系是σ-代数（也叫σ-域）就可以了。σ-代数的基本定义是：1. 全集在里面；2. 里面每个集合的余集在里面；3. 里面任意可列个集合的并集在里面。有了这三条基本定义，就可以推出：空集、可列次交、并、差、上限集、下限集运算之后都能在里面。就满足需要了。

所以，集合X+该集合上的一个σ-代数F，（X，F）就是一个可测空间了，即可以定义测度的空间（F中任一集合都可以定义其测度（某种度量））。进一步再定义了测度μ，那么（X, F, μ）就是测度空间。

对应到概率论中，样本空间Ω，事件域F（是个σ-代数），概率测度P，放一起（Ω，F，P）就是概率测度空间。概率测度P是满足特殊要求的一种测度：$P(\Omega)=1$.

Borel Feild就是Borel σ-代数，表示实数轴上的σ-代数，可由实轴上的所有开集生成（的σ-代数），也可由实数轴上所有的$(-\infty,a]$这样的区间生成（σ-代数），是相等的。按σ-代数前面说的，实数轴上开集、闭集的至多可列次交、并、差(余)、上限集、下限集、极限集的运算，都超不出该Borel σ-代数的范围。

Borel σ-代数（我用Br表示）有什么用？其实概率论中的随机变量，对应测度论中的可测函数，而可测函数就是从可测空间（X，F)到（R，Br）的可测映射。

实数集

再说说随机变量，前面说了概率论中要用集合表示事件，但事件五花八门，怎么统一用一种简单的集合表示呢？这就用到映射的概念，建立一种从样本空间（基本事件的全体）到实数轴的映射（一一对应）就可以了，这种映射就是随机变量。有了它，基本事件映射到实数轴上就是的基本区间，基本事件经过运算生成的复杂事件，映射到实数轴上就是实数轴上Borel σ-代数中的集合。

因为有了这个对应关系，要度量“事件发生的可能性的大小”（即概率测度），只要度量“实数轴上Borel σ-代数中的集合” 就可以了（前面说了Br因为是σ-代数是可以定义测度的，给Br中的集合定义概率测度就行了）。

所以，随机变量的测度论语言定义是这样的：设（Ω，F，P）为概率测度空间，若对实数轴上Borel σ-代数中的任一集合（称为Borel集）B，都有 {w∈Ω: X(w) ∈B} ∈F，则称X(w)为随机变量。

总之，随机变量就是建立了“随机事件”到“实数轴上Borel σ-代数”的一种对应，并且保证了建立了这种对应的随机事件都是可以定义概率测度的。

既然随机事件{w∈Ω: x(w) ∈B}属于F，那么可以有概率，即P{w∈Ω: x(w) ∈B}是有意义的，为了简单，概率中就记P{w∈Ω: X(w)∈B} = P{X ∈B} 了。

特别地，若取B=(-∞,x), 则事件{X∈B}的概率

P{X∈B} = P{X≤x} := F(x)

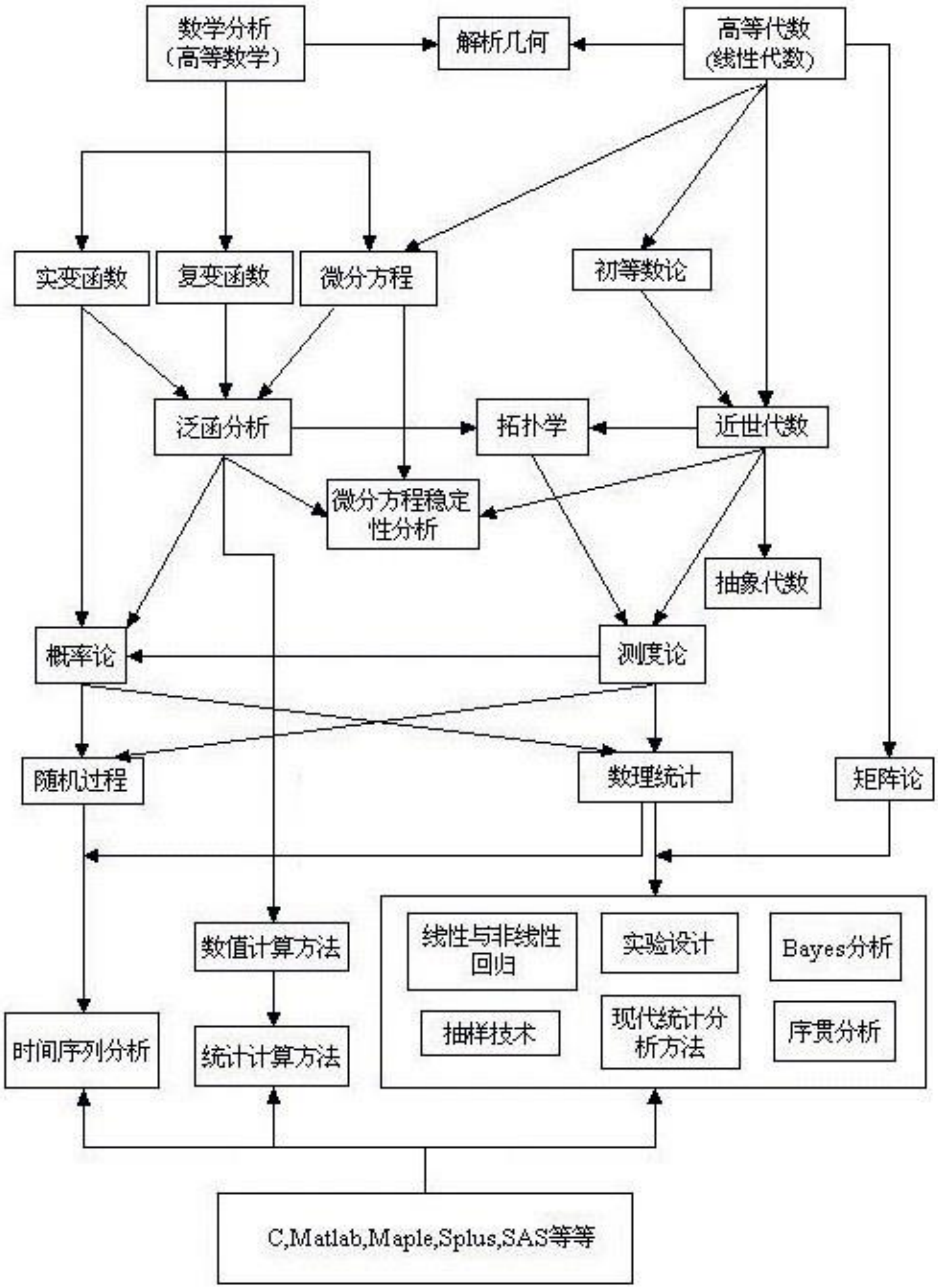
就定义成随机变量X的分布函数。因为对任意的区间(a,b], 都可表示成

P{X ∈(a,b] } = P{a<X≤b} = P{X≤b} - P{X≤a} = F(b)-F(a)

进而，由这样的区间经过至多可列次交、并、差运算的复杂的实数轴上的Borel集都可以用F(x)给出其概率。

当然，随机变量也可以定义为从样本空间到平面R2上的映射，就是二元随机变量。

(d)


数学分析
（高等数学）
解析几何
高等代数
(线性代数)
实变函数
复变函数
微分方程
初等数论
泛函分析
拓扑学
近世代数
微分方程稳定
性分析
抽象代数
概率论
测度论
随机过程
数理统计
矩阵论
数值计算方法
线性与非线性
回归
实验设计
Bayes分析
时间序列分析
统计计算方法
抽样技术
现代统计分
析方法
序贯分析
C,Matlab,Maple,Splus,SAS等等

(e) https://math.stackexchange.com/questions/2286357/relation-between-events-and-borel-field-in-probability-theory

So probability is defined as a measure $\mathbb{P}$ over some topological space Ω.

The measure $\mathbb{P}$ is a map that maps subsets (events) of Ω into a real number that is between 0 and 1. But it needs to meet some criteria, basically we want:

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

$$\mathbb{P}(\Omega) = 1$$

For events pairwise disjoint $\{A_i\}_{i=1}^{\infty}$

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

In order for the above to be true, we found that usually we cannot define a proper $\mathbb{P}$ for all subsets of Ω, but only for subsets consisting a σ-algebra.

And for probability it's best to define it over the σ-algebra generated by open sets of the topological space Ω. For example, it will be most convenient to keep the measurability for composite of random variables.

(f)

3. 来看看"以P表示测度"是怎么用的（从此开启各种分布函数之旅）

举个栗子：

对于事件A "随机变量 X 的值为正"， 可以表示为这样一个样本点集合：

$\{\omega \in \Omega : X(\omega) > 0\}$

把它简写为 $\{X>0\}$, 则这个事件对应的概率P(A) = P($\{X>0\}$) 我们经常也会简写为 P(X>0)

那么 P(X>0) 这个东东实质上就是集合 $\{\omega \in \Omega : X(\omega) > 0\}$ 的测度。（或者说，它实质上是一个函数，这个函数的值为{ } 的测度。）

眼熟？眼熟就对了。分布函数就是这么定义的。

看我大课本肿么说：

Def：

称 $F(x) = P\{ X(\omega) < x\}$, $-\infty < x < +\infty$

为随机变量X(ω) 的分布函数

...........公理化结构搭建起来了之后就有无限的可能..........

(g) 引言）测度论基础——从古典型概率到现代概率 (测度论基础（0）)
https://zhuanlan.zhihu.com/p/27776420

Kolmogorov 是现代公理化概率论的奠基人，自他开始，基于测度的概率分析方式得到越来越多的关注，也使得我们能用的工具越来越强大，对这个世界的客观描述方式越来越丰富化。在正式的介绍概率的测度论基础之前，我想先聊一聊为什么要引入概率测度，以及古典型概率的局限性。

概率这个东西，既可以看作是我们对某一事件发生的可能性的一种 belief，也可以看作是重复多次某一事件结果可能出现的频率的描述。比如我们说扔一枚 fair 的硬币出现 head 和 tail 的概率各为 1/2，那么它既可以理解为我们下一次投掷出现这两种情况的机率都是 1/2，也可以理解为我重复很多次投掷，head 和 tail 的结果大概是各占一半。根据我们的常识，我们对于一个事件的概率会有一下几个要求：

(i) 所有可能的结果概率之和为一;
(ii) 每一个可能的结果的概率都是大于等于零的
(iii) 几种不同的结果组合到一起的概率等于这些结果的概率之和

当可能性是有限的时候，这个东西十分好理解的——扔一枚骰子的概率，考试分数的概率等等，但是当我们想要描述的对象变成无限的时候问题就来了，人类对于抽象的无限的东西认知总是没有有限的东西来的深。

我们可以先看看无限但是countable（可数 / 可列）的情况，这个时候我们要借助的工具就是级数 $\sum_{n=1}^{\infty} p_n = 1$（$p_n$ 代表结果是第n种情况的可能）。我们都学过在某些情况下，这无穷多项 p_n 的和是可能为1的 比如 $\sum_{n=1}^{\infty} \frac{1}{2^n}$，当然这样的数列还有很多，可以刻画我们日常生活中遇到的很多问题，比如单位时间内到达银行的人数可以建模成一个Poisson分布。那么对于这类情况我们就有了一个比较好用的数学工具，而且这与我们的常识1，2，3并不违背。

但是当我们想要描述的对象变成无限不可数的时候，问题就变的复杂了。不过还好我们实际当中遇到的大多数不可数集都是cardinality of continuum（简单来说就是实数 R 那么大），我们可以暂时focus 在 R 上。那么问题就来了，在这个R上我们该如何定义概率？我们先来看看我们一直是怎么做的吧！没错，我们每个人应该都听说过的一种分布——正态分布，就是这样的一种定义在实数上的一个分布。如果你学过基础的微积分和概率论的话，应该对这个过程很熟悉了:我们有一个density function $f(x)$，一个值落在区间 $(a, b]$ 上的概率就是 $\int_a^b f(x)dx$。那么我们可以假设，对于任意一个给定的 R 上的集合 X，都有 $P(X) = \int_X f(x)dx$。这个时候我们就遇到了第一个问题，对于形状不那么规则的X，这样的积分怎么定义？进一步的，我们也会好奇是不是所有的X都有这么一个积分存在？会不会有什么意想不到的情况出现？

我们先来看一个比较极端的例子来说明我们的担心是有必要的。我们考虑一维空间上[0,1]集合上的length measure，我们将[0,1]区间按照等价关系划分成这样一些集合：

1. 我们记x~y($x, y \in [0,1]$) iff(if and only if) $x - y \in Q$ (有理数)
2. 从每个等价类当中选取一个元素组成集合计作X
3. 对于任意的 $r \in Q$ ，记 $X_r = \{(N+r) \cup (N+r-1)\} \cap [0,1]$ （即将X右移r个单位，然后将超出了[0,1]的部分挪回来）

那么关于这些 X_r 我们可以验证：

1. 任意的 $a \in [0,1]$ ，都存在一个唯一的 X_r 使得 $a \in X_r$ ----> $\sum_{r \in Q} L(X_r) = 1$
2. 这些 X_r 是disjoint的 ----> 任意的 $L(X_r)$ 都是相等的

那么我们就可以看看这里我们会得到什么样的结果了：

- 如果 $L(X_r) = 0$ ，那么这与1的结论矛盾
- 另外的，如果 $L(X_r) = b > 0$,那么这也与1矛盾

感觉整个世界都不好了！不管怎么样，我们都得到了矛盾，而且看来这个矛盾是没办法解决的，好像只有两种情况啊！等于0不行，不等于0也不行！再仔细想想，其实还有一种情况——那就是我们构造出来的这样的X不在我们所能讨论的范围之内。

回顾整个过程，我们从我们的经验常识想要推广一个概率的概念从有限（finite）到无限（countable），再到更复杂的无限（cardiality of contiuum），与此同时，我们想保证一些基本的性质1），2），3）不被违背。这其实就是数学公里化的一种思想——我们建立一个体系，制定一些基本的规则，想要这些规则有用，研究这个系统对于我们实际当中可能遇到的情况能有所帮助。但是，我们会遇到一些苦难的情况，比如我们上述构造的 $\{X_r\}$ ，这种情况我们没有办法处理——这就涉及到了公理系统的第二个原则，自洽性，就是自己和自己不能相矛盾的。

那么关于概率研究，我们的任务也就非常简单了——构造一个自洽并有用的系统，来帮助我们理解一些事情。实际上，不管是有用的或者没用的，只要你能说明为什么要研究这个，你就可以构造任何你想要的公理系统，并按照这个发展一系列的理论，这个是每个个体的权利——你可以把这个当成玩具来玩！但是要真正做些有用有意义的东西就要走很长的路了。

回到我们的概率问题上来，我们要做的就是定义一个有用的系统，告诉我们什么样的集合是可以有概率的，这个概率是怎么样定义的，这就是概率的测度论基础想要做的事情。当然测度论本身就是一个有很多人研究的东西，它本身也有很多的发展，在这里我只会介绍一点简单的之后会用到的东西，如果对这个感兴趣网上也有很多的教程可以参考。

下一章我会正式介绍测度的东西，more mathematics more fun！

zhangsan 1年前

L是什么呀

赞 回复 踩 举报

韩光 回复 zhangsan 5个月前

我的理解是概率质量函数,测度,或者任何什么从X_r到R的映射,让所有的L和为1

赞

(h) 测度论基础（一）(https://zhuanlan.zhihu.com/p/27827507)

上一次讲到了对于一个给定的空间，我们有可能并不能研究其上的任意一个集合的measure，这一章我们就来从数学的角度严格定义一下什么叫做measure，更具体的，我们会讨论概率measrue。

basic definition

首先我们来定义一下什么样的集合是可测的（measurable），再来定义其上的measure。当然我们也可以先定义measure，然后再将其上的所有可测集刻画出来，形成我们想要的研究对象。我们先考虑第一种，再来考虑第二种。

我们首先来回忆一下上一章中我们将到的概率可能需要满足的三个条件，我们会想讨论两个集合的交，并，以及一个集合的补的概率，那么我们很自然的一种想法就是定义一个由全集 X 上的子集构成的collection $\mathcal{A}$,要求其满足：

1. 如果 $A \in \mathcal{A}$ ，那么它的补 $A^c \in \mathcal{A}$
2. a)如果对于有限个集合 $\{A_i\}_{I=1}^n$ ，如果 $A_1, \cdots, A_n \in \mathcal{A}$ ，那么 $\cup_{I=1}^n A_i \in \mathcal{A}$
3. 由de Morgan'law $A \cap B = (A^c \cup B^c)^c$ ，我们可以得知2）中这些集合的交也属于 $\mathcal{A}$

看起来我们好像得到了我们想要的东西，但是实际上我们还far from satisfied——想想我们在微积分中经常会见到的一个形式 $(a,b] = \cap_{n=1}^{\infty}(a, b+1/n)$,可数个开区间的交或者可数个闭区间的并，我们并不想lose this basic generality in analysis，实际上这种形式的countable intersection\union也是我们在之后的研究中经常用到的，它大大的丰富了我们的研究对象，尤其是考虑contimuous的相关性质的时候。那么，我们可以将上述中的2）改成

1. (same to 1 above)
2. b)如果对于任意countable的 $\{A_i\}_{i=1}^{\infty}$,如果所有的 $A_i \in \mathcal{A}$ ，那么 $\cup_{I=1}^{\infty} A_i \in \mathcal{A}$

满足1），2.a）的collection我们称之为一个algebra，满足1），2.b）的collection我们称之为 $\sigma - algebra$ (也有人会用field来代替algebra，personally I prefer to use algebra)。

定义好了这个底层的集合的collection以后，自然的我们就可以定义其上的measure了（for generality we donot specify it as a probability now）。定义一个从 $\sigma - algebra \quad \mathcal{A}$ 到非负数 R^+ 上的函数 $m : \mathcal{A} \to R^+$ ，它满足以下条件：

1. $\forall A \in \mathcal{A}, m(A) \geq 0$
2. $\forall \{A_i\}_1^{\infty}$,如果 $A_i \in \mathcal{A}$ 且disjoint，那么我们有 $m(\cup_1^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i)$

那么这个triple $(X, \mathcal{A}, m)$ 就被称为一个measure，$(X, \mathcal{A})$ 被称为一个measurable space。如果我们有 $m(X) = 1$,那么这就是一个 probability space。到这里好像一切都结束了，我们有了一系列的集合，定义在这个集合之上的一个measure，它们可以满足我们第一章中所要求的那三条性质。然而，再仔细想想，我们除了知道一些空洞的定义以外，其实什么都不知道。So why not go furthur to discuss more?

Property

对于上述定义的一个measure，我们可以首先看一看它的一些基本的性质

1. $\forall A, B \in \mathcal{A}, \text{if } A \subset B, \text{then } m(A) \leq m(B)$
2. $\forall \{A_n\} \in \mathcal{A}, A_n \subset A_{n+1}$, then $A = \cup_{i=1}^{n} A_i \in \mathcal{A}, \lim_{n \to \infty} m(A_n) = m(A)$
3. $\forall \{A_n\} \in \mathcal{A}, A_n \supset A_{n+1}$, and $m(A_1) < \infty$ ， then $A = \lim_{n \to \infty} \cap_{i=1}^{n} A_i \in \mathcal{A}, \lim_{n \to \infty} m(A_n) = m(A)$

在验证这些的时候，我们需要用到一个技巧

将原来的集合disjoint化。之所以要强调这一点是因为此时我们讨论的是在我们的基本定义之下的一个公理系统，所有的fact就只有之前的那几条，我们能讨论的只有countable disjoint的集合的measure，其它的所有东西都是需要验证的，需要在这之上系统的发展出来，不能根据我们的直觉就拿出来用的。

第一条很好验证，第二条也可以先disjoint化以后再来看，利用之前的条件2很好验证，验证第三条时我们可以首先构造 $B_n = A_1/A_n$ ，这样我们就得到了满足第二条中条件的集合序列，然后根据第二条中的结论可以得到 $m(A_1) - m(A) = m(A_1) - \lim_{n\to\infty} m(A_n)$ ，此时我们就可要用到假设 $m(A_1) < \infty$ 了（其实只要任意一个 A_n 满足这一条件就好了），否则我们并不能得到我们想要的结果。反例可以考虑在自然数集合 N 上的count measure： $m(I) = |I|$ (集合I的measure就是它其中元素的个数)。令 $A_n = \{n, n+1, \cdots\}$ （n到无穷），那么每一个 A_n 的measure都是 ∞ ，但是它们的intersection是 $\emptyset$,measure为0.

第二条和第三条是我们最基本的特殊符号的互换——极限和measure度量的互换，这也是我们后续积分和极限的互换的基础。

这一次就先写到这里吧，下一次我们重点来讲讲怎么样构造一个满足我们想要的性质 $\sigma - algebra$ 以及它之上定义的measure，这个也是一个很有意思的过程！

(i) 测度论基础（2） (https://zhuanlan.zhihu.com/p/27964702)

今天我们接着讲讲概率的测度论基础，上一次我们讲到了algebra， $\sigma - algebra$ ，以及 $(X, \mathcal{A}, m)$ 这样的measure triple，这次我们先举几个具体的例子，再来讨论如何构造这样的triple。

Examples

1. 考虑全体非负整数 $X = \{0, 1, 2, 3, \cdots\}$ ，定义其上的 $\sigma - algebra = 2^X$ （即X的全体幂集），那么它上面的一个counting measure 即可定义为 $m(A) = |A|, \forall A \in \mathcal{A}$ （A中元素的个数），我们很容易来check 它满足我们之前对于measure的要求。
2. 若是对于一个uncountable的set X我们可以这样定义其上的 $\sigma - algebra : \mathcal{A} = \{A \subset X | A \text{ is countable or } A^c \text{ is countable}\}$ ，而m可以定义成这样： $m(A) = 0 \text{ if A is countable}; m(A) = 1 \text{ otherwise}$.第一眼看上去好像有点不太对头，不过我们只要弄清楚一点就行了对于co-countable的集合（即补集为countable的情况），我们无法找到可数个不相交的co-countable的集合（否则它们的补集会构成整个X，而countable 个countable set的union仍然是countable的，这里就有矛盾了）。

这里就很自然的我们想回到我们最开始引入测度的时候的例子了，我们如何定义一个与实数 R 上的length相合的measure?

先来看看我们的要求：

1. 对于给定的集合 $[a,b](a,b),(a,b],[a,b)(a<b)$ ，我们希望它属于 $\mathcal{A}$,且它们的measure均 等于 $m([a,b])=b-a$
2. 可数个这样的集合的并也属于 $\mathcal{A}$ ，且有相应的measure

这其实是有点头疼的——对于 $\sigma-algebra$ ，我们只要求它满足什么样的性质，并没有告诉我们该怎么样找到这样的集合让它包含我们想要的集合。这个时候我们就要用到数学里面经常用的一种定义方式，其实也很有意思——我们可以构造满足我们想要性质的最小的集合——$\sigma-algebra$ generated by $\mathcal{S}$,这里的 $\mathcal{S}$ 是所有的 R 上的开区间所组成的集合。构造它的方式是这样：假设我们现在有所有的包含 $\mathcal{S}$ 的 $\sigma-algebra$ （可以知道这样的 $\sigma-algebra$ 是存在的，至少有R的幂集 2^R ），我们很容易验证这些 $\sigma-algebra$ 的交也是一个 $\sigma-algebra$ ，即为包含 $\mathcal{S}$ 的最小的一个 $\sigma-algebra$ ，我们也将其计作 $\sigma(\mathcal{S})$.

> Remark：1. R 上的所有开区间都是由countable disjoint (a_i,b_i) 的区间interval组成的
> 2.可以验证 $[a,b](a,b),(a,b],[a,b)(a<b)$ ，都是在 $\sigma(\mathcal{S})$ 中的。例如 $(a,b]=\cap_{n=1}^{\infty}(a,b+1/n)$

我们把这个由 R 上所有开区间 $\mathcal{S}$ generated 的 $\sigma(\mathcal{S})$ 叫做R上的Borel-algebra，计作 $\mathcal{B}_R$.

我们姑且认为这个就是我们想要的那个 $\sigma-algebra$ ，好了现在我们来定义在上面measure，这个时候我们又回到了我们之前提出的问题，我们还是无法准确的知道这个 $\sigma-algebra$ 长什么样子，$\mathcal{B}_R$ 只是我们认为的构造出来的一个十分抽象的集合（至少我们无法恨直接的列举出来这个集合中的所有元素的形式）。这个时候我们又要用到一个十分常用的技巧了——用我们能确定的东西来逼近我们不能理解的抽象的东西。下一次我们会介绍outer measure，pre-measure,以及Caratheodory's thoerem,这就十分抽象了，到时候我们继续探讨。

(j) 测度论基础（3） (https://zhuanlan.zhihu.com/p/29968667)

今天我们接着我们上一次的内容继续写， 如何得到一个满足一些基本条件measure （$(X, \mathcal{A}, m)$），比如说定义在 $\mathcal{R}$ 上的与线段长度相一致的measure。

首先我们需要给出一个outer measure的概念：

> $\mu^* : \mathcal{P}(X) \to [0, \infty]$ 从X的幂集到非负轴的函数背成为outer measure，如果它满足一下三个条件：
> 1. $\mu^*(\emptyset) = 0$
> 2. $\mu^*(A) \le \mu^*(B), \text{if } A \subset B$
> 3. $\mu^*(\cup_1^\infty A_j) \le \sum_1^\infty \mu^*(A_j)$

相比于measure，outer measure的要求有强化也有弱化：

1. 它要求作用域是X的所有密集
2. 它不要求disjoint可列可加性，只要求一个相对较弱的subadditivity性质

这个概念乍一看比较奇怪，不知道它会有什么用，不过我们可以确定的是所有的measure都是outer measure。

再回到我们最开始的目的上，假设我们有一系列基本集合以及在其上的一些函数值，那么我们就可以利用它们来构造一个outer measure：

> Proposition：设 $\mathcal{E} \subset \mathcal{P}(X)$ 是 包涵空集和全集的 X 的子集的集合， $\rho : \mathcal{E} \to [0, \infty]$ 满足 $\rho(\emptyset) = 0$ 那么
> $\mu^*(A) = \inf\{\sum_1^\infty \rho(E_j) : E_j \in \mathcal{E} \text{ and } A \subset \cup_1^\infty E_j\}$
> 定义了一个outer measure。

[lower bound of a set]
集合
inf：下确界：Infimum
Ex： $A = \{1, 18, \frac{1}{2}, 7\}$
$\inf\{A\} = \frac{1}{2}$

第一条和第二条很容易验证，第三条我们需要利用inf的性质，给定 ϵ ,对于所有的 A_j 存在一系列 $\{E_{ij}\}$ 使得 $A_j \subset \cup_i E_{ij}$ 满足 $\sum_{i=1}^\infty \rho(E_{ij}) \le \mu^*(A_j) + 2^{-j}\epsilon$ ，那么我们可以得到 $\mu^*(\cup A_j) \le \sum_{i,j} \rho(E_{ij}) + \epsilon$ ，再根据 ϵ 的任意性，我们可以得到outer measure的第三条性质。

有了outer measure μ^* 以后，我们需要再定义一个看起来更加不直接的概念

> 一个集合 A 被称为 μ^* -measurable如果它满足对于任意 $E \subset X$ 都有
> $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$

我们再次借用二维平面这个比较直观的例子来说明一些问题，我们可以看看图形的面积怎么和我们这里所提出的抽象概念相切合。这里也想说一下我个人认为的在数学学习中的具体和抽象的关系——对于初学者来说，具体例子肯定是有好处的，它可以加强自己在直观上的印象，以及为什么要研究这个；但是到了一定的阶段我们就开始要慢慢的忘记这些具体的例子，从公理系统中最基本的假设出发，一步步往前，因为我们能想到的例子往往都太过于平凡，可能会有一些比我们研究假设更强的性质，这个时候我们反而可能需要记住一些特殊的有用的范例（比如那本著名的《实分析中的反例》）。

在二维平面上我们比较了解的一种基本图形就是矩形，对于矩形的面积也十分好定义，而二维平面全集也可以看作是一个面积无穷大的矩形，那么我们就有了之前定义的 $\mathcal{E}$ 和 ρ ，也可以定义相应的outer measure。那么这时，对于二维平面上任意的一个有界的图形A，如果我们取一个包含该图形的大矩形E，那么此时A是outer measurable的条件就是

> $\mu^*(E) - \mu^*(E/A) = \mu^*(A)$

我们可以回想一下，右边就是我们我们在微积分当中学到过的用无限多的外接小矩形来近似图形的外面积，左边就是一个图形的内面积，而当时我们说一个图形的面积是存在的条件也是它的外面积等于它的内面积。通过这个具体的例子，我们大致能够了解这一套系统是怎么工作的。当然outer measurable实际上比内面积等于外面积要来得强一些。

下面我们就咬降到在measure理论中第一个非常重要的定理了——Caratheodory's Theory.

> 如果 μ^* 是一个X上的outer measure，那么所有 μ^* -measurable的集合所构成的collection $\mathcal{M}$ 是一个 $\sigma - measure$,且在 $\mathcal{M}$ 上 μ^* 是一个完备的measure。

这样我们就完成了我们最开始想要做的事情了——得到一个满足一些基本条件measure 。具体的例子和一些相关的讨论我们就下次再写了！

03/08/2020

1. “条件概率分布”: Wiki

条件概率分布（Conditional Probability Distribution，或者 **条件分布**，Conditional Distribution ）是现代概率论中的概念。已知两个相关的随机变量*X* 和*Y*，随机变量*Y* 在条件{*X* =x}下的条件概率分布是指当已知*X* 的取值为某个特定值*x*之时，*Y* 的概率分布。 如果*Y* 在条件{*X* =x}下的条件概率分布是连续分布，那么其密度函数称作*Y* 在条件{*X* =x}下的**条件概率密度函数**（**条件分布密度**、**条件密度函数**）。与条件分布有关的概念，常常以“条件”作为前缀，如条件期望、条件方差等等。

假设在桌子上抛掷一枚普通的骰子，则其点数结果的概率分布是集合$\{1,2,3,4,5,6\}$的均匀分布：每个点数出现的概率都是均等的六分之一。然而，如果据某个坐在桌边的人观察，向着他的侧面是6点，那么，在此条件下，向上的一面不可能是6点，也不可能是6点对面的1点。因此，在此条件下，抛骰子的点数结果是集合$\{2,3,4,5\}$的均匀分布：有四分之一的可能性出现$2,3,4,5$四种点数中的一种。可以看出，增加的条件或信息量（某个侧面是6点）导致了点数结果的概率分布的变化。这个新的概率分布就是条件概率分布。

连续条件分布 [编辑]

对于连续型的随机变量*X*和*Y*，$P(X=i)=P(Y=j)=0$，因此对离散型随机变量的条件分布定义不适用。假设其联合密度函数为$f(x,y)$，*X*和*Y*的边际密度函数分别是$f_X(x)$和$f_Y(y)$，那么*Y*在条件{*X* =x}下的条件概率密度函数是：

$$f_{Y|X}(y|x)=f_Y(y\mid X=x)=\frac{f(x,y)}{f_X(x)}.$$

同样的，*X*在条件{*Y*=y}下的条件概率密度函数是：

$$f_{X|Y}(x|y)=f_X(x\mid Y=y)=\frac{f(x,y)}{f_Y(y)}.$$

条件分布和独立分布 [编辑]

在一定意义上，条件分布和独立分布是相对的。如果两个随机变量*X*和*Y*是独立分布的，那么不论是否已知某个关于*X*的条件，都不会影响*Y*的概率分布。用数学语言来说，就是：

$$P(Y=y\mid X=x)=P(Y=y)=p_Y(y)$$

这与独立分布的定义是相合的，事实上，随机变量*X*和*Y*相互独立分布，则：

$$P(Y=y,X=x)=P(Y=y)\cdot P(X=x).$$

因此

$$P(Y=y)=\frac{P(Y=y,X=x)}{P(X=x)}=P(Y=y\mid X=x).$$

2) “条件期望”: Wiki

“在概率论中，**条件期望是一个实数随机变量的相对于一个条件概率分布的期望值**。换句话说，这是给定的一个或多个其他变量的值一个变量的期望值。它也被称为条件期望值或条件均值。条件期望的概念在柯尔莫哥洛夫(**Kolmogorov**)的**测度理论概率论**的定义很重要。条件概率的概念是由条件期望来定义。”

计算 [编辑]

设X和Y是离散随机变量，则X在给定事件$Y=y$条件时的条件期望是x的在Y的值域的函数

$$\mathrm{E}(X|Y=y)=\sum_{x\in\mathcal{X}}x\ \mathrm{P}(X=x|Y=y)=\sum_{x\in\mathcal{X}}x\ \frac{\mathrm{P}(X=x,Y=y)}{\mathrm{P}(Y=y)},$$

其中，$\mathcal{X}$是处于X的值域。

如果现在X是一个连续随机变量，而Y仍然是一个离散变量，条件期望是：

$$\mathrm{E}(X|Y=y)=\int_{\mathcal{X}}xf_X(x|Y=y)dx$$

其中，$f_X(\cdot|Y=y)$是在给定$Y=y$下X的条件概率密度函数。

3) “鞅 (概率论) (**Martingale**)”: Wiki

在概率论中，鞅（英语：martingale）是满足下述条件的随机过程：已知过去某一 时刻 s 以及之前所有时刻的观测值，若某一时刻 t 的观测值的条件期望等于过去某一时刻 s 的观测值，则称这一随机过程是鞅。而于博弈论中，鞅经常用来作为公平博弈的数学模型。

定义 [编辑]

离散时间鞅 [编辑]

离散时间鞅是对于所有 n 都满足

$$\mathbf{E}(|X_n|)<\infty$$

$$\mathbf{E}(X_{n+1}\mid X_1,\ldots,X_n)=X_n,\quad n\in\mathbb{N},$$

的时间离散的随机过程 X_1，X_2，X_3，…，也就是说，已知之前所有观测值，若下一次观测值的条件期望等于本次观测值，则称这一随机过程（即随机变量序列）是离散时间鞅。

关于随机过程的离散时间鞅 [编辑]

相对来说更为一般的定义如下：若对于所有 n 都满足

$$\mathbf{E}(|Y_n|)<\infty$$

$$\mathbf{E}(Y_{n+1}\mid X_1,\ldots,X_n)=Y_n,\quad n\in\mathbb{N},$$

则称随机过程 Y_1，Y_2，Y_3，…是关于另一随机过程 X_1，X_2，X_3，…的鞅。

连续时间鞅 [编辑]

与离散时间鞅的定义相似，连续时 间鞅的定义为：若对于所有 t 都满足

$$\mathbf{E}(|Y_t|)<\infty$$

$$\mathbf{E}(Y_t\mid\{X_\tau,\tau\le s\})=Y_s,\ \forall\, s\le t,$$

则称关于随机过程 X_t 的连续时间鞅是随机过程 Y_t 。

上述定义表达了鞅的性质，即在 s ≤ t 的条件下，已知时刻 s 以及之前所有时刻的观测值，若时刻 t 的观测值的条件期望等于时刻 s 的观测值，则随机过程是鞅。

“上鞅”: 通常是有递减的趋势。

“下鞅”: 通常是有递增的趋势。
这里给出一个区分下鞅和上鞅的记忆方法：“生活是一个上鞅：随着时间的推进，期望降低。”

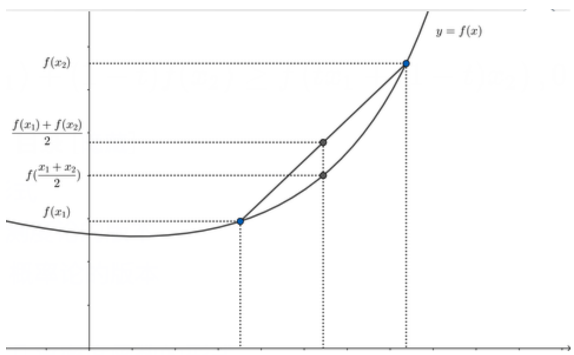
4) “詹森（Jensen）不等式”: Wiki，又叫 “琴生不等式” 或者 “延森不等式”

琴生不等式 [编辑]

维基百科，自由的百科全书

琴生不等式（Jensen's inequality）以丹麦数学家约翰·琴生（Johan Jensen）命名。它给出积分的凸函数值和凸函数的积分值间的关系。琴生不等式有以下推论：过一个凸函数上任意两点所作割线一定在这两点间的函数图象的上方，即：

$$tf(x_1)+(1-t)f(x_2)\geq f(tx_1+(1-t)x_2),0\leq t\leq 1.$$



凸函数是具有如下特性的一个定义在某个向量空间的凸子集C（区间）上的实值函数f：对其定义域C上的任意两点x_1,x_2，总有$f(\frac{x_1+x_2}{2})\leq\frac{f(x_1)+f(x_2)}{2}$。

一般形式 [编辑]

琴生不等式可以用测度论或概率论的语言给出。这两种方式都表明同一个很一般的结果。

测度论的版本 [编辑]

假设μ是集合Ω的正测度，使得$\mu(\Omega)=1$。若g是勒贝格可积的实值函数，而φ是在g的值域上定义的凸函数，则

$$\varphi\left(\int_\Omega g\,d\mu\right)\leq\int_\Omega\varphi\circ g\,d\mu。$$

概率论的版本 [编辑]

以概率论的名词，μ是个概率测度。函数g换作实值随机变数X（就纯数学而言，两者没有分别）。在Ω空间上，任何函数相对于概率测度μ的积分就成了期望值。这不等式就说，若φ是任一凸函数，则

$$\varphi\left(E(X)\right)\leq E(\varphi(X))。$$

4）

代表一个样本点

2.1 有限概率空间

有限概率空间被用于随机试验的所有可能结果有限这种情形的建模。在上一章二叉树模型中，我们有限次抛掷硬币。如果抛掷三次，所有可能结果的集合为：

$$\Omega=\{HHH,HHT,HTH,HTT,THH,THT,TTH,TTT\} \tag{2.1.1}$$

假设每次抛掷出现正面的概率（真实概率或者风险中性概率）为 p，出现背面的概率为 $q=1-p$，并假设抛掷过程是相互独立的，则 Ω 中每个元素 ω（即三次抛掷的结果序列 $\omega=\omega_1\omega_2\omega_3$）的概率为：

$$\mathbb{P}(HHH)=p^3,\quad \mathbb{P}(HHT)=p^2q,\quad \mathbb{P}(HTH)=p^2q,\quad \mathbb{P}(HTT)=pq^2,$$
$$\mathbb{P}(THH)=p^2q,\quad \mathbb{P}(THT)=pq^2,\quad \mathbb{P}(TTH)=pq^2,\quad \mathbb{P}(TTT)=q^3 \tag{2.1.2}$$

代表另一个样本点，etc

Ω 中的子集称为**事件**，事件也经常用文字或符号来描述。例如：

Ω 中的子集称为**事件**，事件也经常用文字或符号来描述。例如：

$$\text{“第一次抛掷结果为正面”}=\{\omega\in\Omega;\omega_1=H\}$$
$$=\{HHH,HHT,HTH,HTT\}$$

事件的概率由事件中各元素的概率相加得到，例如：

$$\begin{aligned}\mathbb{P}(\text{第一次抛掷结果为正面})&=\mathbb{P}(HHH)+\mathbb{P}(HHT)\\&\quad+\mathbb{P}(HTH)+\mathbb{P}(HTT)\\&=(p^3+p^2q)+(p^2q+pq^2)\\&=p^2(p+q)+pq(p+q)\\&=p^2+pq\\&=p(p+q)\end{aligned}$$

者。在式(2.1.3)的情形下，我们试图建立一种模型，其中每次抛掷出现正面的概率为 p，为此我们先由式(2.1.2)定义 Ω 中元素的概率，进而定义一个事件的概率为该事件中元素的概率和。这些定义令我们可以施行式(2.1.3)中的计算，我们也需要进行这种计算来验证是否能得到预期的答案。若不然，我们就要重新考虑抛掷硬币的数学模型。

我们将刚才讨论的情形稍加推广，首先允许 Ω 为任意有限集，其次允许 Ω 中某些元素的概率为零。这就导致了下面的定义。

定义 2.1.1 **有限概率空间包含样本空间 Ω 和概率测度 $\mathbb{P}$。** 样本空间是一个非空有限集合，概率测度 $\mathbb{P}$ 为一个函数，该函数将 Ω 中每个元素对应到[0,1]之间的某个数，使得：

$$\sum_{\omega\in\Omega}\mathbb{P}(\omega)=1 \tag{2.1.4}$$

一个事件是 Ω 中的一个子集，我们按如下方式定义事件 A 的概率为：

$$\mathbb{P}(A)=\sum_{\omega\in A}\mathbb{P}(\omega) \tag{2.1.5}$$

如上面提到的，这是一个关于随机试验的模型。集合 Ω 为试验的所有可能结果，$\mathbb{P}(\omega)$ 为某个特别结果 ω 发生的概率，$\mathbb{P}(A)$ 为结果在样本空间的子集 A 中的概率。如果 $\mathbb{P}(A)=0$，那么试验结果一定不在 A 中；如果 $\mathbb{P}(A)=1$，那么试验结果一定在 A 中。由于式(2.1.4)，我们有：

$$\mathbb{P}(\Omega)=1 \tag{2.1.6}$$

也就是说，试验的结果一定在 Ω 中。因为对某些 ω，$\mathbb{P}(\omega)$ 允许为零，我们甚至可以将一些一定不会发生的结果加入 Ω 中。显然，由式(2.1.5)可知，如果 A 和 B 为 Ω 中不相交的子集，那么：

$$\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B) \tag{2.1.7}$$

1.2 Random Variables, Distributions, and Expectations – Shreve I (2.2)

Definition 2.2.1. *Let $(\Omega,\mathbb{P})$ be a finite probability space. A random variable is a real-valued function defined on Ω. (We sometimes also permit a random variable to take the values $+\infty$ and $-\infty$.)*

RV: Ω->R

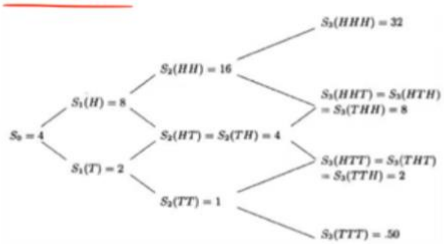


Fig. 1.2.2. A particular three-period model.

2.2 随机变量、分布和期望

随机试验一般会得到一些数据，这就产生了随机变量的概念。

定义 2.2.1 设$(\Omega,\mathbb{P})$为有限概率空间，**随机变量**为定义在Ω上的一个实值函数（有时我们也允许随机变量取值为$+\infty$和$-\infty$）。

例 2.2.2（股票价格） 回忆由式(2.1.1)给出的三次独立抛掷硬币结果

04/09/20

1) Important! (Borel 集的作用？意义？它为什么重要？)

(https://www.zhihu.com/question/33991971)

我来从概率论的角度说一下Borel集的意义吧。

首先回忆一下概率空间的定义，概率空间是一个三元组，样本空间、事件的集合、概率。

我们的概率函数是定义在事件上的，而不是定义在样本空间上的。

概率的定义必须满足几个性质，也就是Kolmogorov公理：

定义 2. （**Kolmogorov axioms**）给定一个样本空间 Ω 以及相应的 σ-代数 $\mathscr{F}$，函数 $\mathscr{P}:\mathscr{F}\to[0,1]$ 若满足：

1. 对于所有的事件 $A\in\mathscr{F}$，$\mathscr{P}(A)\geq 0$
2. $\mathscr{P}(\Omega)=1$
3. 若 $A_1,A_2,\ldots\in\mathscr{F}$ 为两两互斥事件，则 $\mathscr{P}\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\mathscr{P}(A_i)$（**可数可加性或可列可加性**）

则我们称 $\mathscr{P}$ 为**概率函数**或**概率测度**。

当样本空间离散的时候，概率非常容易定义，只要分配给每一个结果一个概率值，其和等于1就好了。这个时候我们**可以定义概率的所有事件**的集合就是样本空间的全部子集，没毛病。

但是当样本空间不是离散的，比如说是实轴R，那么概率的定义就比较麻烦了。我们发现，**并不是R上的所有子集（事件）都可以被定义上概率**，比如：

例 6. （勒贝格不可测集）如果我们选取样本空间 $\Omega = [0,1]$，令 Ω 中的所有有理数集合为 Q'，由于有理数为可数集合，因而可以写成 $Q' = \{q_1, q_2, \ldots\}$。对于

$(0,1)$ 之间的任意实数 a，定义集合

$$S_a = \left\{ \begin{array}{ccc} a+q & if & a+q<1 \\ a+q-1 & if & a+q \geqslant 1 \end{array} \forall q \in Q' \right\}$$

那么可知 $\bigcup_{a\in(0,1)} S_a = [0,1]$。由于 S_a 也是可数集，因而可以将其写为：

$$S_a = \{s_{a1}, s_{a2}, \ldots\}$$

令 T_1 为所有 S_a 中的 s_{a1}，T_2 为所有 S_a 中的 s_{a2}，因而我们有可数个 T_k，$\bigcup_{k=1}^{\infty} T_k = [0,1]$，且 T_k 两两不相交。每个 T_k 地位相等因而 $\mathscr{P}(T_k) = \mathscr{P}(T_{k'})$。若 $\mathscr{P}(T_k) > 0$，则：

$$1 = \mathscr{P}([0,1]) = \mathscr{P}\left(\bigcup_{k=1}^{\infty} T_k\right) = \sum_{k=1}^{\infty} \mathscr{P}(T_k) = \infty$$

如果 $\mathscr{P}(T_k) = 0$，则：

$$1 = \mathscr{P}([0,1]) = \mathscr{P}\left(\bigcup_{k=1}^{\infty} T_k\right) = \sum_{k=1}^{\infty} \mathscr{P}(T_k) = 0$$

无论如何都会得到矛盾。

看，我们发现，如果严格按照概率函数的定义，并不是R上的所有子集（事件）都能被定义概率的。那么我们能不能把那些能定义概率的集合（事件）挑出来只对他们进行研究呢?

当然可以。但是挑出来的这些事件的集合必须有一些要求，比如：

1. 空集必须在这个事件的集合里面
2. 如果我们关心某个事件A，那么「事件A不发生」，也就是A的补集也必须在这个事件的集合里面
3. 如果我们关心一系列事件，那么这些事件同时发生、至少有一个发生的概率也必须能被研究。

这就诞生了所谓的σ-代数的概念。

那么我们怎么在R上定义概率呢？这个时候我们需要引入分布函数了：

定义 3. （**分布函数**）如果函数 $F:\mathbb{R}\to\mathbb{R}$ 满足：

1. 单调性：$F(a)\leq F(b), a\leq b$
2. 右连续：$\lim_{x\to x_0^+}F(x)=F(x_0)$
3. $F(-\infty)=0, F(\infty)=1$

有了分布函数，我们可以先在区间上定义概率：

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

则对于任意的 $-\infty<a<b<+\infty$，有：

$$\begin{cases} P((a,b])=F(b)-F(a) \\ P((a,b))=F(b-)-F(a) \\ P([a,b))=F(b-)-F(a-) \\ P([a,b])=F(b)-F(a-) \end{cases}$$

然后通过所谓的内测度外测度定义出一些集合的概率函数。R上所有的区间所生成的最小σ-代数就是Borel σ-代数，Borel σ-代数的元素叫做Borel集。我们发现，刚好所有的Borel集都是可以被定义上概率的，而且是最符合我们直觉的（除非特意构造，碰到的多数集合都是Borel集），所以我们就干脆只研究Borel集的概率算了。

所以它为什么重要？因为通过它，我们排除了那些不好的集合，限制了我们讨论的范围，把我们的问题简单化了。

dddd

1.2 “Ito” 积分