# Lecture 1 - Introduction & Preliminaries

赵尉辰

南开大学 统计与数据科学学院

# 目录

1. 课程概况
2. 概率论的公理化体系
3. 随机变量
4. 随机变量的收敛
5. Filtration($\sigma-$代数流)

# 目录


1. 课程概况
2. 概率论的公理化体系
3. 随机变量
4. 随机变量的收敛
5. Filtration($\sigma$−代数流)

## 本人信息

- 授课老师：赵尉辰；
- 电子邮箱：zhaoweichen@nankai.edu.cn；
- 个人主页：
  https://my.nankai.edu.cn/stat/zwc/list.htm
- 研究领域：
  图深度学习方法及应用；
  随机算法: 采样、随机优化、生成式AI；
  深度学习理论.

# 课程信息

- 课程主页：
  https://weichenzhao1996.github.io/WeichenZhao.io/STAT0041-2024.html

- 教材：Lecture Note
  参考书：
  1. 钱忠民，应坚刚 随机分析引论
  2. Bernt Øksendal Stochastic differential equations: an introduction with applications
  3. 龚光鲁 随机微分方程及其应用概要
  4. 黄志远 随机分析学基础（第二版）

- 教学方式：板书为主，Slides为辅

- 考核方式：
  - 5次平时作业 30% (纸质版，latex/手写);
  - 出勤 10% (签到表);
  - 期末考试 60% (5-6道题).

# 课程简介

- 随机分析导论/应用随机分析；
  概率论专业/涉及随机分析工具的交叉学科(统计物理、金融数学、人工智能、随机优化)
- 预备课程：概率论、随机过程、(实分析、泛函分析)
- 随机分析的应用：建模随机现象

  建模粒子的运动：统计物理、化学；
  建模金融产品的价格波动：金融数学；
  建模数据分布的演化：人工智能。

# 课程简介

- (数学/实)分析："好"的函数(光滑/可测)→微积分(黎曼/勒贝格)→微分方程
  随机分析："好"的随机过程→随机微积分→随机微分方程
- 主要内容：
  - 概率论基础
  - 鞅论初步
  - 布朗运动
  - Itô随机积分
  - 随机微分方程
  - 随机分析在人工智能中的应用

# 目录


1. 课程概况
2. 概率论的公理化体系
3. 随机变量
4. 随机变量的收敛
5. Filtration($\sigma-$代数流)

## 样本空间

概率论是研究随机现象确定性规律的理论。为了使用数学工具，我们首先要建立随机现象的数学模型。

在概率论中，我们假定随机试验(random trial)可以在相同条件下重复地进行，每次试验的结果可能不止一个，并且能事先确定试验的所有可能结果，但每次试验的结果事先又不可预测。这样一组定义明确的可能结果，称为样本空间。

**定义 1 (样本空间Sample Space)**

把随机试验每一个可能的结果称为一个样本点 *(sample point)*, 通常用 $\omega$ 表示，所有可能的结果组成的集合称为样本空间*(sample space)*, 通常用 $\Omega$表示.

考虑先后掷两次硬币可能出现的结果是：(正, 正)(正, 反)(反, 正)(反, 反)，把这四个结果作为样本点构成这个随机试验样本空间。

# 事件

事实上，我们感兴趣的是试验中出现的一些事，比如，先后掷两次硬币这个随机试验中我们可能感兴趣 “两次出现的结果相同” 这件事，它是指 (正, 正)(反, 反) 这两个样本点之一出现。这些“事”是样本点的集合，称为事件。

**定义 2 (事件Event)**

**事件** *(event)* 定义为样本点的某个集合. 称某事件发生当且仅当它所包含的某个样本点出现。

我们把样本空间$\Omega$本身也作为一个事件。每次试验必然有$\Omega$中的某个样本点出现，即$\Omega$必然发生。因此，我们称$\Omega$为必然事件(certain event)。我们把空集$\emptyset$也作为一个事件，每次试验中，它都不发生，因此，称为$\emptyset$为不可能事件(impossible event)。

思考：必然事件 vs 发生概率为1的事件

# 事件

我们需要能够用简单的事件来刻画复杂的事件，这是由集合的运算实现的。

- 我们称事件$A$发生意味着事件$B$发生，如果$A \subset B$.
  $A = B \iff A \subset B$ 且 $B \subset A$;
- 由所有不包含在事件$A$中的样本点所组成的事件称为事件$A$的对立事件，记为$A^c$;
- 用$A \cap B$ or $AB$表示$A$和$B$都发生;
- 用$A \cup B$表示事件$A$和$B$至少有一个发生；
- 用$A \backslash B$表示事件$A$发生，但是$B$不发生;

# 事件域

如果我们对事件$A$感兴趣，那么我们应该知道与$A$相关的事件，也就是我们需要找出通过集合运算得到的事件。所有这些事件的集合，称为事件域。

**定义 3 (事件域Event field)**

$\mathscr{F}$是由样本空间$\Omega$的一些子集组成的集合，称为**事件域** *(event field)*如果满足:

(1) 非空 $\mathscr{F} \neq \emptyset$;

(2) 对补运算封闭 $A \in \mathscr{F} \Longrightarrow A^c \in \mathscr{F}$;

(3) 对可列并运算封闭 $A_n \in \mathscr{F}, n = 1, 2, \cdots \Longrightarrow \cup_{n=1}^{\infty} A_n \in \mathscr{F}$.

# 事件域

事件与事件域是紧密联系的，在表述事件时，必须明确是在哪个事件域中。

### Example 1

考虑有两个正方形盒子，被分成四个区域，其中盒子*1*盒盖是完全透明的，盒子*2*的乙和丁区域是不透明的。考虑盒子中均有一个小球可以滚动，随意晃动盒子，小球停止运动后，随机地停留在这四块区域中的某块中*(*假设处于理想状态，不考虑小球处于分割线*)*。

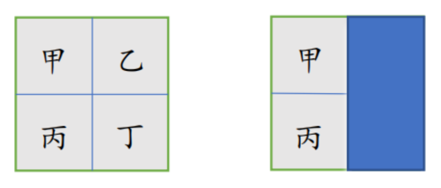

图 1　　图 2

# 事件域

这两个随机试验的样本空间均为$\Omega = \{甲,乙,丙,丁\}$。

但是试验1的事件域$\mathscr{F}_1$为$\Omega$的全体子集组成的集合，其中共有 16 个事件。对于试验1的每一个结果$\omega$，对于$\Omega$的每个子集$A$，总能判断出$\omega$是否属于$A$，也就是说每次实验后，总能知道事件$A$是否发生。

试验2的事件域

$$\mathscr{F}_2 = \{\Omega, \emptyset, \{甲\}, \{丙\}, \{甲，丙\}, \{乙，丁\}, \{甲，乙，丁\}, \{丙，乙，丁\}\},$$

只包含了 8 个事件。对于试验的某些结果，虽然可以看到小球停留在不透明的区域，但是我们不能判断此时小球是否在“乙”区域，也就是说，不知道 $\{乙\}$是否发生了. 因此，对于事件域$\mathscr{F}_2$，$\Omega$的子集$\{乙\}$就不是事件。同样的，$\{丁\}$ 和 $\{甲，乙\}$ 等也不是事件。

## 概率

搞清楚了我们感兴趣的事件的集合，我们现在需要量化它发生的可能性，这就需要引入概率。

### 定义 4 (概率Probability measure)

称集合函数$\mathrm{P}$是事件域$\mathscr{F}$上的概率*(Probability measure)*，如果

*(1)* 非负性：$\mathrm{P}(A) \geq 0, \forall A \in \mathscr{F}$;

*(2)* $\mathrm{P}(\emptyset) = 0,\ \mathrm{P}(\Omega) = 1$;

*(3)* 可列可加性: 对于不相交的集合$A_n \in \mathscr{F}, n = 1, 2, \cdots$

$$\mathrm{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

## 概率空间

总结一下，我们需要三个部分来建立随机现象的数学模型：

- 随机试验的样本空间$\Omega$，它是包含了这个试验所有可能结果的非空集合；
- 事件域$\mathscr{F}$，它是我们感兴趣的事件，以及这些事件通过运算得到的事件的全体；
- 概率$\mathrm{P}$，它量化了事件发生的可能性。

**定义 5 (概率空间Probability space)**

概率空间*(Probability space)*是一个数学三元组$(\Omega, \mathscr{F}, \mathrm{P})$.

# 补充概念

在测度论中，满足事件域定义的集合族$\mathscr{F}$也称为$\sigma$-代数($\sigma$-algebra)或者$\sigma$-域。$(\Omega, \mathscr{F})$称为一个可测空间(measurable space)。可以定义$\mathscr{F}$上的集合函数$\mu$满足非负性和可列可加性，称为$\mathscr{F}$上的测度(measure)，$(\Omega, \mathscr{F}, \mu)$称为测度空间。概率空间是一类特殊的测度空间。

生成$\sigma$-代数：设$\mathscr{C}$是$\Omega$的非空集族，称$\mathscr{S}$是$\mathscr{C}$生成的$\sigma$-代数，如果
(1) $\mathscr{C} \subset \mathscr{S}$；(2) 对任意$\Omega$上的$\sigma$-代数$\tilde{\mathscr{S}}$，如果$\mathscr{C} \subset \tilde{\mathscr{S}}$，那么$\mathscr{S} \subset \tilde{\mathscr{S}}$。
即$\mathscr{S}$是包含$\mathscr{C}$的最小$\sigma$-代数，记为$\sigma(\mathscr{C})$。对任意$\mathscr{C}$，$\sigma(\mathscr{C})$是存在唯一的。

我们介绍$\mathbb{R}$上一类常见的$\sigma$-代数：Borel $\sigma$-代数.
记$\mathbb{R}$上左开右闭区间组成的集合族为

$$\mathscr{C} = \{(a, b] : a, b \in \mathbb{R}, a < b\}.$$

称$\sigma(\mathscr{C})$为$\mathbb{R}$上的Borel $\sigma$-代数，记为$\mathscr{B}(\mathbb{R})$，其中的元素称为Borel集。事实上，对于左闭右开区间组成的集合族、开区间组成的集合族以及闭区间组成的集合族都生成$\mathscr{B}(\mathbb{R})$.

# 目录


1. 课程概况
2. 概率论的公理化体系
3. 随机变量
4. 随机变量的收敛
5. Filtration($\sigma-$代数流)

# 随机变量

现实中，随机现象纷繁复杂，相应的样本空间千差万别。有些可以用数来表示，比如测量误差，有些则不行，比如掷一枚硬币。我们希望能将这些试验结果用“数”来表示，最简单的办法就是把样本点映射到一个数上，即$\xi:\Omega\to\mathbb{R}$。但是我们知道，描述事件时需要明确所对应的事件域，这引入了随机变量的概念。

**定义 6 (随机变量Random variable)**

给定概率空间$(\Omega,\mathscr{F},\mathrm{P})$, 随机变量 *(random variable, r.v.)* 是一个函数$\xi:\Omega\to\mathbb{R}$ 满足：对于所有的$B\in\mathscr{B}(\mathbb{R})$, $\{\omega:\xi(\omega)\in B\}\in\mathscr{F}$.

测度论中，这样的函数称为可测函数(measurable function)。

## 基本概念

**定义 7 (概率分布Probability distribution)**

设$\xi$ 是概率空间 $(\Omega, \mathscr{F}, \mathrm{P})$上的随机变量，$\mathscr{B}(\mathbb{R})$上的概率测度

$$\mathrm{P}_\xi : \mathscr{B}(\mathbb{R}) \to [0,1], \quad \mathrm{P}_\xi(A) := \mathrm{P} \circ \xi^{-1}(A), \ \forall A \in \mathscr{B}(\mathbb{R}).$$

称为$\xi$的*概率分布(probability distribution)*.

若两个随机变量$\xi$ 和 $\eta$具有相同的概率分布，则称$\xi$和$\eta$是同分布的。$\xi$ 和 $\eta$可以是两个不同概率空间上的随机变量，但它们可以有相同的分布。

$F(x) = \mathrm{P}(\omega : \xi(\omega) \leq x), x \in \mathbb{R}$称为$\xi$的分布函数(distribution function).

# 数学期望

**定义 8 (Expectation)**

设$\xi$是概率空间$(\Omega, \mathscr{F}, \mathrm{P})$上的随机变量，如果 $\int_\Omega |\xi(\omega)|\mathrm{dP}(\omega) < \infty$，则称$\xi$的数学期望存在，定义

$$\mathbb{E}[\xi] = \int_\Omega \xi(\omega)\mathrm{dP}(\omega) = \int_{\mathbb{R}} x\mathrm{dP}_\xi(x)$$

为$\xi$的数学期望*(mathematical expectation)*.

$\int_\Omega \xi(\omega)\mathrm{dP}(\omega)$为Lebesgue积分。直观上，$\xi(\omega)$的值落入$x$的$\epsilon-$邻域$[x-\epsilon, x+\epsilon)$的概率为$\mathrm{P}(\omega : \xi(\omega) \in [x-\epsilon, x+\epsilon))$，近似作为$\xi(\omega)$在$x$取值的权重。将$\xi(\omega)$的值域划分成之多可列个这样互不相交的区间$[x_i-\epsilon, x_i+\epsilon)$，那么，$\xi(\omega)$的加权平均就近似为

$$\sum_i x_i \mathsf{P}\big(\omega : \xi(\omega) \in [x_i-\varepsilon, x_i+\varepsilon)\big).$$

令$\epsilon \to 0$，上式的极限就是Lebesgue积分$\int_\Omega \xi(\omega)\mathrm{dP}(\omega)$.

# 目录

1 课程概况

2 概率论的公理化体系

3 随机变量

4 随机变量的收敛

5 Filtration($\sigma-$代数流)

# 随机变量的收敛

### 定义 9

设$(\Omega, \mathscr{F}, \mathrm{P})$为一个概率空间, $(X_n)$为$(\Omega, \mathscr{F}, \mathrm{P})$上随机变量序列。

*(1)* 依概率收敛*Convergence in probability:* 记为$X_n \xrightarrow{\mathrm{P}} X$, 如果对于$\epsilon > 0$

$$\mathrm{P}(\omega : |X_n(\omega) - X(\omega)| > \epsilon) \to 0, n \to \infty.$$

*(2)* 几乎处处收敛*Almost sure convergence:* 记为$X_n \to X$, $a.s.$, 如果

$$\mathrm{P}(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1.$$

*(3)* 依分布收敛*Convergence in distribution:* 记为$X_n \xrightarrow{d} X$，如果对于任意有界连续函数$f$

$$\int f \mathrm{dP}_{X_n} \to \int f \mathrm{dP}_X,$$

*(4)* $L^p$收敛*Convergence in* $L^p$*:* 记为$X_n \xrightarrow{L^p} X$, 如果

$$\|X_n - X\|_p = (\mathbb{E}|X_n - X|^p)^{1/p} \to 0.$$

# 随机变量的收敛

$$
\begin{array}{ccccc}
\xrightarrow{L^r} & \underset{r>s\geq 1}{\Rightarrow} & \xrightarrow{L^s} & & \\
 & & \Downarrow & & \\
\xrightarrow{a.s.} & \Rightarrow & \xrightarrow{P} & \Rightarrow & \xrightarrow{d}
\end{array}
$$

**命题 1**

$X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{\mathrm{P}} X$

Proof. 由切比雪夫不等式，对于任意 $\epsilon > 0$ 和$p \geq 1$，有

$$\mathrm{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p}$$

# 目录


1. 课程概况
2. 概率论的公理化体系
3. 随机变量
4. 随机变量的收敛
5. Filtration($\sigma-$代数流)

# Filtration($\sigma-$代数流)

由于不确定性，人们不能精确预测未来，但人们总是希望通过已知过去和现在的信息来预测未来。如何在概率空间的框架下来精确定义这种“信息”？这需要引入Filtration($\sigma-$代数流)。

### 定义 10 (Filtration)

$(\Omega,\mathscr{F},P)$上的 *filtration* 是一族$\mathscr{F}$的子$\sigma-$代数$(\mathscr{F}_t)$由$\mathbb{R}^+\cup\{0,\infty\}/\mathbb{Z}^+\cup\{0,\infty\}$索引，满足 $\mathscr{F}_s\subset\mathscr{F}_t,\ \forall s\leq t\leq\infty$. 定义$\mathscr{F}_\infty:=\sigma\left(\bigcup_n\mathscr{F}_n\right)\subset\mathscr{F}$. 称$(\Omega,\mathscr{F},(\mathscr{F}_t),P)$为一个 *filtered* 概率空间。

# Filtration

## Example 2

考虑先后掷三次硬币这一随机过程，样本空间为

$$\Omega = \{000, 001, 010, 011, 111, 110, 101, 100\} = \{0, 1\}^3$$

样本点表示成$\omega = (\omega_1 \omega_2 \omega_3)$，事件域为$\mathcal{F} = \mathscr{P}(\Omega)$.

掷0次：在掷硬币之前，我们只能确定所有的样本点是什么，但并不知道哪个样本点将出现。我们所能了解的信息仅仅是必然事件$\Omega$发生，不可能事件$\emptyset$不发生，也就是说，这时我们能知道的事件域是$\mathscr{F}_0 = \{\emptyset, \Omega\}$。

# Filtration

掷1次：虽然试验还未完成，我们不能预测具体某个样本点$\omega$是否最终出现，但这时已经知道$\omega$的部分 “信息”。

若掷1次得到的结果是$\omega_1 = 1$，那么我们知道事件“第一次是正面”发生，事件“第一次是反面”不发生，加上知道的必然事件和不可能事件，我们知道以下四个事件

$$
\begin{aligned}
\omega \in A_1 = \{\text{第一次是正面}\} = \{111, 110, 101, 100\}, \\
\omega \notin A_0 = \{\text{第一次是反面}\} = \{000, 001, 010, 011\}, \\
\omega \in \Omega, \\
\omega \notin \emptyset.
\end{aligned}
$$

同理，若掷1次得到的结果是$\omega_1 = 0$，我们知道事件“第一次是正面”不发生，而事件“第一次是反面”发生，还知道必然事件和不可能事件，所以这时我们知道的事件域

$$\mathscr{F}_1 = \{\emptyset, A_1, A_0, \Omega\}$$

# Filtration

掷2次：当第1次和第2次试验完成，若结果是$\omega_1\omega_2 = 10$，那么我们知道以下六个事件的信息

$$\omega \in A_{10} = \{100, 101\}, \omega \notin A_{11} = \{110, 111\}, \omega \in \Omega,$$
$$\omega \notin A_{00} = \{000, 001\}, \omega \notin A_{01} = \{010, 011\}, \omega \notin \emptyset.$$

此外，我们知道$\omega$是否属于这几个事件的交、并、对立事件，以及其交、并、对立事件再交、并和对立事件，即我们此时知道的事件域

$$\mathscr{F}_2 = \{\Omega, \emptyset, A_1, A_0, A_{11}, A_{10}, A_{01}, A_{00}, A_{11}^c, A_{10}^c, A_{01}^c, A_{00}^c, A_{11} \cup A_{01}, A_{11} \cup A_{00}, A_{10} \cup A_{01}, A_{01} \cup A_{00}\}.$$

掷3次：当3次试验都完成后，我们知道$\mathscr{F}_3 := \mathscr{F}$中所有事件的信息，即，对于任何$A \in \mathscr{F}_3$，我们知道$A$是否发生。

# Filtration

**定义 11 (适应过程)**

一个随机过程$(X_t)$称为$\mathscr{F}_t$-适应过程, 如果对于任意$t$, $X_t$是$\mathscr{F}_t$-可测的。

**定义 12 (通常条件Usual condition)**

我们称*filtered*概率空间 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathrm{P})$满足通常条件，如果

*(1)* 右连续性*:* $\forall t \geq 0, \mathscr{F}_t = \mathscr{F}_{t+} := \bigcap_{\delta\downarrow 0} \mathscr{F}_{t+\delta}$.

*(2)* 完备性*:* $\mathscr{F}_0$ *contains all* $\mathrm{P}$*-null set.*

**Example 3 (自然$\sigma$-代数流Natural filtration)**

设$X = (X_t)$ 为一随机过程, $X$的自然$\sigma$-代数流 *natural filtration*定义为

$$\mathscr{F}_t^X = \sigma(X_s, 0 \leq s \leq t), \quad \mathscr{F}_\infty^X = \sigma(X_s, s \geq 0).$$

*Natural filtration*是使得$X$适应的最小 *filtration.*

# 总结

- 样本空间：包含了随机试验所有可能结果的集合；
- 事件域：事件、以及这些事件通过运算得到的事件的集合(集族)；
- 概率：事件域上的函数，量化事件发生的可能性；
- 概率空间$(\Omega, \mathscr{F}, \mathrm{P})$：建立了随机性的数学模型
- 随机变量：样本空间上的实值可测($\forall B \in \mathscr{B}(\mathbb{R}), \{\omega : \xi(\omega) \in B\} \in \mathscr{F}$)函数；
- Filtration：递增的子事件域集合，刻画了过去和现在已知的信息。