Lecture 1 - Introduction & Preliminaries

赵尉辰

南开大学 统计与数据科学学院

目录

目录

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、扩散模型；
  深度学习理论.

课程信息

- 课程主页:
  https://weichenzhao1996.github.io/WeichenZhao.io/STAT0008-2025.html
  智慧小雅

- 教材：Sheldon M. Ross, Stochastic Processes. 2nd Edition
  Sheldon M. Ross 著，龚光鲁 译，随机过程(第2版)
  参考书:
  1. James R. Norris, Markov chains.
  2. 陆大絟 张颢，随机过程及其应用(第2版)

- 教学方式：板书+Slides

- 考核方式:
  - 5次平时作业 30% (纸质版，latex/手写);
  - 出勤 10% (签到);
  - 期末考试 60% (出7-8道题，选取得分最高的5道题记录成绩).

课程简介

- 随机过程导论/应用随机过程;
- 预备课程：初等概率论、(实分析、泛函分析)
- 主要内容：
  - 概率论基础
  - Poisson过程
  - Markov链
  - 连续时间Markov链 (排队论)
  - Markov过程* (状态连续)
  - 随机过程在数据科学中的应用 (MCMC算法, 模拟退火, Markov决策过程*)
- 后续课程：随机分析

目录

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 ω 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 Ω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 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Ω本身也作为一个事件。每次试验必然有Ω中的某个样本点出现，即Ω必然发生。因此, 我们称Ω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)。

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Ω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。

例 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Ω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ω，对于Ω的每个子集A，总能判断出ω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$\mathscr{F}_2$，Ω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P是事件域$\mathscr{F}$上的概率*(Probability measure)*，如果

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

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

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

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

概率的基本性质

- 若 $E \subset F$, 则 $P(E) \leq P(F)$. (单调性 monotonicity)
- $P(E^c) = 1 - P(E)$.
- $P(\bigcup_i E_i) \leq \sum_i P(E_i)$ (次可加性 subadditivity/ Boole's inequality)

概率的连续性

- 对于一个递增事件序列 $\{E_n, n \geq 1\}$，即，$E_1 \subset E_2 \subset \cdots$，定义其极限事件为

$$\lim_{n\to\infty} E_n \triangleq \bigcup_{i=1}^{\infty} E_i.$$

- 对于一个递减事件序列 $\{E_n, n \geq 1\}$，即，$E_1 \supset E_2 \supset \cdots$，定义其极限事件为

$$\lim_{n\to\infty} E_n \triangleq \bigcap_{i=1}^{\infty} E_i.$$

命题 1

如果$\{E_n, n \geq 1\}$是一个递增或者递减序列，那么

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

Borel-Cantelli 引理

设$E_1, E_2, \ldots$是一个事件序列，定义：

$$\limsup_{i\to\infty} E_i \triangleq \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} E_i = \{\text{无穷多个}E_i\text{发生}\} = \{\omega \in \Omega : \forall n, \exists i \geqslant n, \text{使}\omega \in E_i\}$$

若有无穷多个E_i发生，则对每个n，$\bigcup_{i=n}^{\infty} E_i$都发生，从而$\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} E_i$发生。
另一方面，如果事件$\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} E_i$发生，则对所有$n$，$\bigcup_{i=n}^{\infty} E_i$发生，从而对所有$n$，至少有个$i \geq n$，使得$E_i$发生，因此有无穷多个$E_i$发生。

思考.[1] $\liminf$情形：

$$\liminf_{i\to\infty} E_i = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} E_i = \{\text{至多有限多个}E_i\text{不发生}\} = \{\omega \in \Omega : \exists n, \forall i \geqslant n, \omega \in E_i\}$$

[1] 更多集合论的内容参考：严加安，《测度论讲义》，1.1节

Borel-Cantelli 引理

命题 2 (Borel-Cantelli 引理)

若 $\sum_{i=1}^{\infty} P(E_i) < \infty$, 那么

$$P(\text{无穷多个}E_i\text{发生}) = 0.$$

Proof. 注意到 $\bigcup_{i=n}^{\infty} E_i, n \geq 1$是递减的事件序列，由概率的连续性：

$$\begin{aligned} P(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i) =& P(\lim_{n\to\infty} \bigcup_{i=n}^{\infty} E_i) = \lim_{n\to\infty} P(\bigcup_{i=n}^{\infty} E_i) \\ &\leq \lim_{n\to\infty} \sum_{i=n}^{\infty} P(E_i) = 0. \end{aligned}$$

最后一个等号成立是因为： $\sum_{i=1}^{\infty} P(E_i) < \infty$. □

Borel-Cantelli 引理

命题 3 (Borel-Cantelli 引理 II)

设$E_1, E_2, \ldots$是一个独立事件序列，若 $\sum_{i=1}^{\infty} P(E_i) = \infty$, 那么

$$P(\text{无穷多个}E_i\text{发生}) = 1.$$

Proof. 对于任意 $n < m < \infty$, 注意到 $1 - x \le e^{-x}$, 于是

$$\begin{aligned} P(\bigcap_{i=n}^{m} E_i^c) &= \prod_{i=n}^{m} P(E_i^c) \qquad\qquad \text{(由独立性)} \\ &= \prod_{i=n}^{m}(1 - P(E_i)) \le \prod_{i=n}^{m} \exp(-P(E_i)) = \exp\left[-\sum_{i=n}^{m} P(E_i)\right] \xrightarrow{m\to\infty} 0. \end{aligned}$$

因此, $P(\cup_{i=n}^{m} E_i) \to 1$ as $m \to \infty$, $P(\cup_{i=n}^{\infty} E_i) = 1$ 对任意n成立。那么

$$P(\text{无穷多个}E_i\text{发生}) = P(\lim_{n\to\infty} \bigcup_{i=n}^{\infty} E_i) = \lim_{n\to\infty} P\left(\bigcup_{i=n}^{\infty} E_i\right) = 1.$$

Borel-Cantelli 引理

例 2

设$X_1, X_2, \ldots$独立且满足

$$P(X_n = 0) = \frac{1}{n},\ P(X_n = 1) = 1 - \frac{1}{n},\ n \geq 1$$

考虑事件$E_n = \{X_n = 0\}$，由于 $\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$，由*Borel-Cantelli* 引理 *II*，

$$P(\text{无穷多个}E_n\text{发生}) = 1.$$

我们是否可以说，以概率1，$\lim_{n\to\infty} X_n = 0$?

答案是否定的，由于$\sum_{n=1}^{\infty} P(E_n^c) = \sum_{n=1}^{\infty} (1 - \frac{1}{n}) = \infty$，由*Borel-Cantelli* 引理 *II*，

$$P(\text{无穷多个}E_n^c\text{发生}) = 1.$$

于是以概率1，$n \to \infty$，X_n没有极限值。

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Ω，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σ-代数(σ-algebra)或者σ-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μ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σ-代数：设$\mathscr{C}$是Ω的非空集族，称$\mathscr{S}$是$\mathscr{C}$生成的σ-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σ-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σ-代数：Borel σ-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 σ-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

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$X : \Omega \to \mathbb{R}$ 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)。

分布

定义 7 (概率分布Probability distribution)

设X 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

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

称为X的*概率分布(probability distribution)*.

若两个随机变量X 和 Y具有相同的概率分布, 则称X和Y是同分布的。X 和 Y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。

分布函数

定义 8 (分布函数)

对于任意$x \in \mathbb{R}$，随机变量X的分布函数*(distribution function)*定义为

$$F(x) = P(X \leq x) = P(\{\omega; X(\omega) \in (-\infty, x]\}), \qquad \bar{F}(x) = 1 - F(x) = P(X > x).$$

如果一个随机变量X可能值的集合是可数的，则称它是离散随机变量，它的分布函数为

$$F(x) = \sum_{y \leq x} P(X = y).$$

如果存在一个函数$f(x)$(称为概率密度函数, probability density function, pdf)，使得对于任意(Borel) 集合 $B \subset \mathbb{R}$

$$P(X \in B) = \int_B f(x) dx,$$

则称随机变量X是连续的。

联合分布

两个随机变量X和Y的联合分布函数定义为

$$F(x,y) = P(X \le x, Y \le y)$$

X和Y的分布函数分别为

$$F_X(x) = P(X \le x) = \lim_{y \to \infty} P(X \le x, Y \le y) = \lim_{y \to \infty} F(x,y).$$
$$F_Y(y) = P(Y \le y) = \lim_{x \to \infty} F(x,y).$$

如果存在一个函数$f(x,y)$，使得对于任何(Borel)集合A和B，有

$$P(X \in A, Y \in B) = \int_A \int_B f(x,y)dydx$$

则称随机变量X和Y是联合连续的(Jointly continuous)，$f(x,y)$称为联合密度函数。

独立性

定义 9 (独立)

如果对于一切x，y，有

$$F(x,y)=F_X(x)F_Y(y)$$

则称随机变量X和Y是独立的。

更一般地，任意一族随机变量$X_1,X_2,\ldots,X_n$的联合分布定义为

$$F(x_1,\ldots,x_n)=P(X_1\le x_1,\ldots,X_n\le x_n).$$

对于$1\le i\le n$，边缘分布为

$$F_{X_i}(x_i)=\lim_{x_j\to\infty,j\ne i}F(x_1,x_2,\ldots,x_n).$$

称$X_1,X_2,\ldots,X_n$是独立的，如果

$$F(x_1,\ldots,x_n)=F_{X_1}(x_1)\cdots F_{X_n}(x_n).$$

目录

期望

定义 10 (期望)

随机变量X的期望$\mathbb{E}[X]$定义为：

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x dF(x) = \begin{cases} \sum_x xP(X = x) & \text{若}X\text{是离散的.} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{若}X\text{是连续的.} \end{cases}$$

如果此积分存在。

对于X的任意函数h，$h(X)$的期望为

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x)dF(x)$$

命题 4

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

方差

定义 11 (方差/协方差)

随机变量X的方差$Var[X]$定义为：

$$Var[X] = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

两个联合地分布的随机变量X和Y的协方差定义为

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

若$Cov(X,Y) = 0$，我们称X和Y不相关(uncorrelated)。所以X和Y独立$\Rightarrow$$X$和$Y$不相关，但反过来不一定正确。

命题 5

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{i<j} Cov(X_i, X_j)$$

组合数学中的概率方法

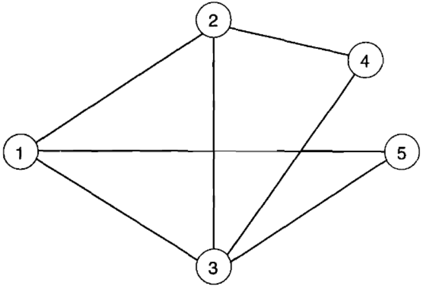
开创概率方法[2]的数学大师：Paul Erdős

[2]感兴趣可参考专著：Alon N, Spencer J H. The probabilistic method[M]. John Wiley & Sons, 2016.

例：组合数学中的概率方法

图(graph) $G=(V,E)$由节点(node)集合V和连边(edge)集合E组成。例如下图中，

$$V=\{1,2,3,4,5\},\quad E=\{(1,2),(1,3),(1,5),(2,3),(2,4),(3,4),(3,5)\}$$



证明：对于任何一个图，一定存在节点集的一个子集A，使得至少有一半的边，它的一个节点在A中，而另一个在A^c中。例如上图中$A=\{1,2,4\}$.

Proof.

转化为数学语言：设一个图有m条边，分别记为$1, 2, \ldots, m$。对于任意节点集合B，恰有一个节点在B中的连边的个数记为$C(B)$，那么我们即是要证明：

$$\max_{B} C(B) \geq \frac{m}{2}.$$

概率建模：随机选取一个节点集合S，使得这个图的任意节点都独立地以概率$1/2$在S中。记随机变量X为：恰有一个节点在S中的连边的个数，它可能的取值是$C(B)$所有可能值的集合。

对于连边$i \in \{1, \ldots, m\}$，如果恰有一个节点在S中，就令$X_i = 1$，否则$X_i = 0$，那么

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{m} X_i\right] = \sum_{i=1}^{m} \mathbb{E}[X_i] = \frac{m}{2}.$$

由于随机变量至少有一个值和它的均值一样大，所以至少有一个节点集合B满足

$$C(B) \geq \frac{m}{2}.$$

□

条件期望

对于离散随机变量 X, Y, 对所有y满足$P(Y = y) > 0$,
给定$Y = y$，X的条件概率质量函数(conditional probability mass function) 定义为

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

给定$Y = y$，X的条件分布函数(Conditional distribution function)定义为

$$F(x|y) = P(X \leq x|Y = y) = \sum_{z \leq x} P(X = z|Y = y).$$

给定$Y = y$，X的条件期望定义为

$$\mathbb{E}[X|Y = y] = \int x dF(x|y) = \sum_{x \in S} xP(X = x|Y = y).$$

条件期望

如果 X, Y 有联合密度函数$f(x,y)$, 对所有y满足$f_Y(y) = \int f(x,y)dx > 0$ ，
给定$Y=y$，X的条件概率密度函数(conditional probability density function) 定义为

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

给定$Y=y$，X的条件分布函数(Conditional distribution function)定义为

$$F(x|y) = \mathbb{P}(X \leq x|Y=y) = \int_{-\infty}^{x} f(z|y)dz.$$

给定$Y=y$，X的条件期望定义为

$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} xf(x|y)dx.$$

条件期望

命题 6

记$\mathbb{E}[X|Y]$为随机变量Y的函数：它在$Y=y$处取值$\mathbb{E}[X|Y=y]$。当天件期望存在时，对于一切随机变量X和Y有

$$\mathbb{E}[X]=\mathbb{E}[\mathbb{E}[X|Y]]=\int \mathbb{E}[X|Y=y]dF_Y(y)$$

注. 初等概率论中的条件概率和条件期望不严格性

- 在初等定义中，要求$P(Y=y)>0$，实际中会遇到概率为0的情况，比如连续空间中单点的概率；
- 初等定义中的条件期望通常基于具体的条件值（$Y=y$），但在更一般的条件下，我们需要定义一个更一般的条件期望，能够处理更复杂的条件信息，如σ-代数。

条件期望与Bayes估计

在Bayes统计中，我们认为观测数据$X=(X_1,\cdots,X_n)$的分布由一个随机变量θ确定，θ服从一个概率分布，我们称为先验分布(prior distribution)。

θ的一个估计$d(X)$是观察数据X的任意函数，Bayes统计中通常选择$d(X)$使得条件均方误差

$$\mathbb{E}[\left(\theta-d(X)\right)^2|X]$$

达到最小。

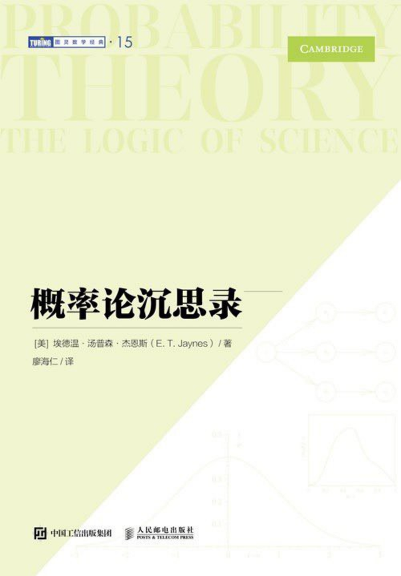
可以证明[3]，使得$\mathbb{E}[\left(\theta-d(X)\right)^2|X]$达到最小的估计为

$$d(X)=\mathbb{E}[\theta|X]$$

称为Bayes估计。

[3]参阅教材1.5节

Bayes观点


PROBABILITY THEORY
THE LOGIC OF SCIENCE
TURING 图灵数学经典 · 15
CAMBRIDGE
概率论沉思录
[美] 埃德温 · 汤普森 · 杰恩斯（E. T. Jaynes）/ 著
廖海仁 / 译
中国工信出版集团
人民邮电出版社
POSTS & TELECOM PRESS

目录

矩母函数

定义 12 (矩母函数)

随机变量X的矩母函数*(moment generating function)*定义为

$$\psi(t) = \mathbb{E}[e^{tX}] = \int e^{tx} dF(x).$$

可以通过对矩母函数ψ求各阶导得到X的各阶矩

$$\psi(0) = 1, \psi'(0) = \mathbb{E}[X], \psi''(0) = \mathbb{E}[X^2], \cdots, \psi^{(n)}(0) = \mathbb{E}[X^n].$$

注. 当矩母函数存在时，它唯一地确定分布。

离散概率分布[4]	概率质量函数 $p(x)$	矩母函数 $\psi(t)$	均值	方差
二项分布 参数 $n, p, 0 \leq p \leq 1$	$\binom{n}{x} p^x(1-p)^{n-x}$ $x = 0, 1, \cdots, n$	$[pe^t + (1-p)]^n$	np	$np(1-p)$
泊松分布 参数 $\lambda > 0$	$e^{-\lambda}\frac{\lambda^x}{x!}$ $x = 0, 1, 2, \cdots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
几何分布 参数 $0 \leq p \leq 1$	$p(1-p)^{x-1}$ $x = 1, 2, \cdots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布 参数 r, p	$\binom{x-1}{r-1} p^r(1-p)^{x-r}$ $x = r, r+1, \cdots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

[4]连续概率分布参考教材表1.4.2

特征函数

由于随机变量的矩母函数可能不存在，注意到$|e^{itX}|=1$，在理论上考虑下面特征函数更为方便。

定义 13

随机变量X的特征函数(*characteristic function*)定义为

$$\varphi(t)=\mathbb{E}[e^{itX}].$$

更一般地，随机变量$X_1,X_2,\ldots,X_n$的联合矩母函数为

$$\psi(t_1,\cdots,t_n)=\mathbb{E}[\exp(\sum_{i=1}^n t_iX_i)],$$

以及联合特征函数为

$$\varphi(t_1,\cdots,t_n)=\mathbb{E}[\exp(i\sum_{i=1}^n t_iX_i)].$$

Markov不等式

引理 1 (Markov不等式)

设X是一个非负随机变量，那么对于任意$a > 0$，有

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

推论 1 (切比雪夫不等式)

设X是一个随机变量，且均值μ和方差σ^2有限，则对任何$k > 0$，有

$$P\{|X - \mu| \geqslant k\} \leqslant \frac{\sigma^2}{k^2}$$

Chernoff界

命题 7 (Chernoff 界)

设X是一个随机变量，其矩母函数为$\psi(t) = \mathbb{E}[e^{tX}]$, 那么对于$a > 0$,

$$\mathbb{P}(X \geq a) \leq e^{-ta}\psi(t), \ \textit{for } t > 0;$$

$$\mathbb{P}(X \leq a) \leq e^{-ta}\psi(t), \ \textit{for } t < 0.$$

Proof.

对于 $t > 0$,

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \overset{\text{Markov}}{\leq} \mathbb{E}[e^{tX}]e^{-ta}.$$

$t < 0$ 同理可证。 □

Jensen不等式

命题 8 (Jensen不等式)

如果 f 是一个凸函数，即

$$\lambda f(x)+(1-\lambda)f(y)\geq f(\lambda x+(1-\lambda)y),\forall\lambda\in(0,1),x,y\in\mathbb{R}.$$

那么只要期望存在，就有

$$\mathbb{E}[f(X)]\geq f\left(\mathbb{E}[X]\right)$$

$f(x)=|x|$是凸的，于是

$$|\mathbb{E}[X]|\leq\mathbb{E}[|X|]$$

$f(x)=x^2$是凸的，于是

$$(\mathbb{E}[X])^2\leq\mathbb{E}[X^2]$$

随机变量的收敛

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

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

极限定理

定理 1 (弱大数定律)

设 $X_1, X_2, \ldots$ 独立同分布且公共期望$\mu = \mathbb{E}[X_i] < \infty$, 则对任意$\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right) = 0.$$ *convergence in probability*

定理 2 (强大数定律)

设$X_1, X_2, \ldots$独立同分布且公共期望$\mu = \mathbb{E}[X_i] < \infty$, 则

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1.$$ *a.s. convergence*

定理 3 (中心极限定理)

设 $X_1, X_2, \ldots$ 独立同分布，且均值为μ，方差为 σ^2, 则对任意 a,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$ *convergence in distribution*

目录

基本概念

固定t

定义 14 (随机过程)

设$(\Omega, \mathscr{F}, \mathrm{P})$为概率空间，$(E, \mathscr{E})$为可测空间，指标集 $T \subset \mathbb{R}$，若对任何 $t \in T$，

$$X_t : \Omega \mapsto E,$$

可测，则称$\{X_t : t \in T\}$是$(\Omega, \mathscr{F}, \mathrm{P})$上的取值于$E$的随机过程，称$(E, \mathscr{E})$为其“相空间”或“状态空间”，称$T$为其“时间域”。

简单来说，随机过程$\{X_t(\omega) : t \in T\}$是一族随机变量，若指标集$T$是可数集，则我们称$\{X_t\}$为离散时间的随机过程，若$T$是连续统，则称$\{X_t\}$为连续时间的随机过程。

基本概念

固定w

定义 15 (样本轨道 Sample path)

设$\{X_t : t \in T\}$是一个取值于E的随机过程。X 的样本轨道定义为在固定$\omega \in \Omega$情况下的映射

$$T \ni t \mapsto X_t(\omega)$$

也就是说，X的样本轨道是那些由$\omega \in \Omega$索引的，从时间集合T到状态空间E的映射的集合

$$\{t \mapsto X_\omega(t)\}_{\omega \in \Omega}$$

本课程内容

- 离散时间离散状态的随机过程：Markov链；
- 连续时间离散状态的随机过程：跳过程 (Poisson过程、连续时间Markov链)；
- 连续时间连续状态的随机过程：Markov过程。