

# Part III-B: Probability Theory and Mathematical Statistics

Lecture by 李漫漫

Note by THF

2024 年 11 月 27 日

## 目录

<b>1 第一章</b>	<b>3</b>
1.1 随机事件 . . . . .	3
1.1.1 现象 . . . . .	3
1.1.2 随机试验 . . . . .	4
1.1.3 样本 . . . . .	4
1.1.4 随机事件 . . . . .	5
1.2 事件关系与运算 . . . . .	5
1.3 事件的概率 . . . . .	5
1.3.1 古典概型 . . . . .	6
1.3.2 几何概型 . . . . .	7
1.4 公理化 . . . . .	9
1.5 条件概率与乘法公式 . . . . .	11
1.6 全概率公式 . . . . .	11
1.7 贝叶斯公式 . . . . .	11
1.8 独立性 . . . . .	13
<b>2 一维随机变量及其分布</b>	<b>14</b>
2.1 随机变量及其分布函数 . . . . .	14
2.2 常见分布律 . . . . .	17
2.3 连续型随机变量 . . . . .	19
2.4 标准化 . . . . .	20
2.5 随机变量函数的分布 . . . . .	21

<b>3</b>	<b>多维随机变量函数及其分布</b>	<b>23</b>
3.1	二维随机变量及其分布	23
3.2	二维随机变量的边缘分布函数	24
3.3	联合分布律	24
3.4	二维连续性随机变量及其概率特性	25
3.5	多维随机变量及分布	27
3.5.1	多维随机变量的独立性	28
3.5.2	条件分布	29
3.6	二维随机变量函数的分布	29
3.7	二元正态分布	31
<b>4</b>	<b>数字特征</b>	<b>33</b>
4.1	数学期望	33
4.2	数学期望的性质	34
4.3	方差的性质	36
4.4	协方差的性质	39
4.5	相关系数	40
4.5.1	标准化	40
4.5.2	性质	41
<b>5</b>	<b>大数定律和中心极限定理</b>	<b>42</b>
5.1	大数定律	43
5.2	中心极限定理	43
<b>6</b>	<b>数理统计基本概念</b>	<b>44</b>
6.1	经验分布函数	45
6.2	密度函数	46
6.3	统计量	47
6.4	样本均值的分布	48
6.4.1	三大抽样分布	49
<b>7</b>	<b>参数估计</b>	<b>50</b>
7.1	矩估计	50
7.2	极大似然估计	51
7.3	区间估计	51
<b>8</b>	<b>假设检验</b>	<b>52</b>

## Lecture 1

09.03

### 概述

#### 资源

公众号: 狗熊会、大数据文摘, 好玩的数学

MOOC: 爱课程, Coursera, Edx, 网易公开课等

#### 教师要求

教材: 概率论与数理统计第二版

参考: The Lady Tasting Tea, 程序员数学之概率统计, ...

学习目的: 自问自答, 自言自语

考核及成绩组成:

期中 (10)

作业与考勤 (10)

期末 (70)

MOOC (10)

#### 课程简介

概率: Probability

统计: Statistics

概率论与数理统计: Probability theory and Mathematical statistics

**Notation.** 第一章重要但不突出

从概率到概率论: 新增时间 (随机事件、样本空间变化)

从统计到数理统计: 统计最开始为记录性质, 后来衍生出预测, 通过数学模型引入数理统计  
类似的还有政府统计、经济统计等

2000-2015 年间, IT 时代逐渐转换为 DT(Data Technology) 时代, 大数据逐渐占时代主体

## 概率论部分

### 1 第一章

#### 1.1 随机事件

##### 1.1.1 现象

确定性现象: 一定条件下必然发生

随机现象强调统计规律性

**Notation.** 统计规律性:

1. 每次试验前不能预测结果
2. 结果不止一个
3. 大量试验下有一定规律

**Example.** 星际旅行时宇航员看到的现象不是随机现象:

对星际旅行的人而言, 无法完成大量试验

宇航员观测到的结果无规律, 只能称为不确定现象 (Uncertain)

**Example.** 扔一个骰子不能预测结果, 但可以知道结果是 1, 2, 3, 4, 5, 6 的一个, 因此观察扔骰子是随机现象 (Random)

### 1.1.2 随机试验

随机试验 (E): 研究随机现象时进行的实验或观察等

**Notation.** 随机试验的特性:

1. 可以在完全相同的条件下重复进行
2. 试验的可能结果在试验前已知
3. 试验的结果不可预测

### 1.1.3 样本

在随机试验中, 不可再分的最简单结果成为样本点  $\omega$ , 全体样本点组成样本空间  $\Omega$

**Notation.** 随机事件是基本事件的集合

**Example.** 扔骰子存在 6 个基本事件, 可以产生  $2^6$  个随机事件

其中样本空间  $\Omega = \{x | x \in [1, 6], x \in \mathbb{R}\}$

**Example.** 1. 射击时用  $\omega_i$  表示击中  $i$  环, 样本空间为:

$$\Omega = \{\omega_0, \omega_1, \omega_2, \dots, \omega_{10}\}.$$

2. 微信用户每天收到信息条数的取值范围是  $[0, +\infty)$ , 样本空间为无限集:

$$\Omega = \{N | N \geq 0, N \in \mathbb{R}\}.$$

3. 电视机的寿命样本空间为  $\Omega = \{t | t > 0\}$ , 为连续的非负实数集

4. 投掷两枚硬币, 样本空间为  $\Omega = \{(x, y) | x, y = 0, 1\}$ , 其中 0, 1 分别代表正面和背面

**Notation.** 1. 样本点可以不是数

2. 样本空间可以是无限集

### 1.1.4 随机事件

## 1.2 事件关系与运算

1.1.  $A \subset B$ :  $A$  发生必然  $B$  发生

1.2.  $A = B$ :  $A \subset B, B \subset A$

2.  $A \cup B$ :  $A$  和  $B$  至少有一个发生

2.1  $A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^n A_i$

3.  $A \cap B$ :  $A$  和  $B$  只发生一个

4.1.  $A, B$  互斥: 不能同时发生:  $AB = \emptyset$

4.2.  $A, B$  对立: 非此即彼:  $A \cup B = \Omega$

5.  $A - B$ :  $A\bar{B}$  或  $A(\Omega - B)$ , 或  $A$  发生但  $B$  不发生

**Notation.**  $A - B = A\bar{B} \subset A$ ,  $B - A = B\bar{A} \subset B$

当  $AB = \emptyset$  时,  $A - B = A, B - A = B$

**Notation.**  $P(\Omega) = 1, P(\emptyset) = 0$ , 且  $P(\Omega) + P(\emptyset) = 1$ , 即  $\Omega$  与  $\emptyset$  互斥

6. 结合律:  $(A \cup B) \cup C = A \cup (B \cup C)$

7. 分配律:  $(AB) \cup C = (A \cup C)(B \cup C)$ ,  $(A \cup B)C = AC \cup BC$

8. 交换律:  $A \cup B = B \cup A, AB = BA$

**Notation.** 德摩根律:

$$\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i.$$

$$\overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \bar{A}_i.$$

**Example.**

$$\overline{A \cup B} = \bar{A}\bar{B}.$$

$$\overline{(A \cup B) \cup C} = \overline{A \cup B \cup C} = \dots$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}.$$

## 1.3 事件的概率

概率分类:

$$\left\{ \begin{array}{l} \text{主观概率} \\ \text{统计概率} \\ \text{古典概型} \\ \text{几何概型} \end{array} \right\}.$$

**Notation.** 德摩根、蒲丰、皮尔逊、维纳均进行过投掷硬币的试验，随着试验次数的增加，出现正面的频率逐渐接近 0.5

大数定律说明，该事件的概率为 0.5

**Definition.** 统计概率： $A$  为试验  $E$  的一个事件，随着重复次数  $n$  的增加， $A$  的频率接近于某个常数  $p$ ，定义事件  $A$  的概率为  $p$ ，记为  $P(A) = p$

频率的特性：

1. 非负性： $f_n(A) \in [0, 1]$
2. 规范性： $f_n(\Omega) = 1$
3. 有限可加性： $A_i$  两两互斥，则  $f_n(\sum_{i=1}^n A_i) = \sum_{i=1}^n f_n(A_i)$

**Definition.** 主观概率：人对某个事件发生与否的可能性的估计

**Definition.** 完备事件组： $A_1, A_2, \dots, A_n$  两两互斥，且

$$P\left(\sum_{i=1}^n A_i\right) = 1.$$

或

$$\sum_{i=1}^n A_i = \Omega.$$

则称  $A_1 \rightarrow A_n$  为完备事件组（不重不漏）

**Example.**  $A, \bar{A}$  是完备事件组

### 1.3.1 古典概型

古典概型特点：有限等可能性（基本事件数有限，基本事件发生的可能性相等）

**Notation.** 概率计算：

$$P(A) = \frac{m}{n} = \frac{n(A)}{n(\Omega)}.$$

**Example.** 某年级有 6 人在 9 月份出生，求 6 个人中没有人同一天过生日的概率

基本事件总数： $30^6$

目标事件： $30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 = P_{30}^6$

概率：

$$P(A) = \frac{P_{30}^6}{30^6}.$$

**Example.** 有  $N$  个乒乓球中有  $M$  个白球、 $N - M$  个白球，任取  $n(n < N)$  个球，分有放回和不放回，求取到  $m$  个黄球的概率

1. 不放回：

基本事件总数： $C_N^n$

目标事件:  $C_M^m C_{N-M}^{n-m}$

概率:

$$P = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, n = \max\{0, n - (N - M)\}, \dots, \min\{n, M\}.$$

2. 有放回:

$$P = \frac{C_n^m M^m (N - M)^{n-m}}{N^n} = C_n^m \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m}, m \in [0, n].$$

注意到该概率为伯努利分布  $C_n^m B(n, \frac{M}{N})$

匹配问题:

**Example.** 麦克斯韦-玻尔兹曼统计问题:

$n$  个质点随机落入  $N$  ( $N > n$ ) 个盒子, 盒子容量不限, 设  $A$  表示指定的  $n$  个盒子各有一个质点,  $B$  表示恰好有  $n$  个盒子装一个质点

基本事件总数:  $N^n$

$A$  考虑顺序, 即:

$$P(A) = \frac{n!}{N^n}.$$

同理:

$$P(B) = \frac{C_N^n}{N^n}.$$

### 1.3.2 几何概型

几何概型特点: 使用事件所对应的**几何度量**计算

$$P(A) = \frac{m(A)}{m(\Omega)}.$$

**Notation.** 度量: 面积、体积、长度等描述几何量大小的测度方式

**Example.** 地面铺满 2 dm 的地砖, 向地面投掷一个  $r = 0.5$  dm 的光盘, 求光盘不与边线相交的概率

如图: 课后习题: A 组 8 题, B 组 3 题

**Example.** 两人相约 8-9 点间在某地相见, 先到的人等待 20 分钟后离去, 求二人会面的概率

设  $(x, y)$  分别表示两人到达的时刻

设  $G$  为样本空间, 绘制样本空间:

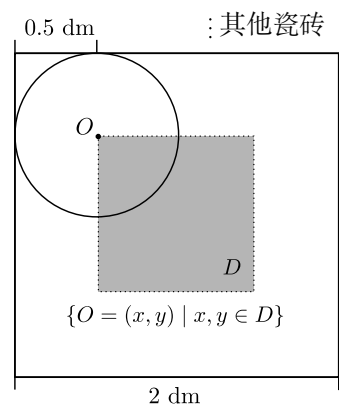
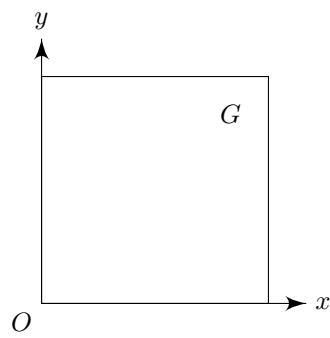


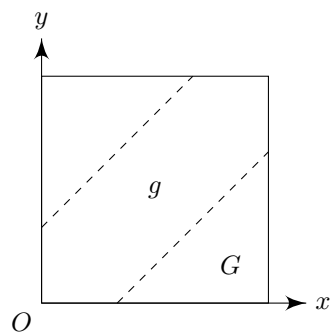
图 1: 例题 1



由题：两人到达的时间之差的绝对值小于 20 分钟 ( $\frac{1}{3}$  小时)，即：

$$|x - y| \leq \frac{1}{3}.$$

将事件绘制：



$$P(g) = \frac{m(g)}{m(G)} = \frac{S(g)}{S(G)} = \frac{1 - \left(\frac{2}{3}\right)^2}{1} = \frac{4}{9}.$$



**Notation.** 几何概型的特点:

1. 非负性:

$$P(A) \in [0, 1].$$

2. 规范性:

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

3. 可列可加性:

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

## Lecture 2

### 1.4 公理化

$$(\Omega, \mathcal{F}, p).$$

**Definition.**  $\Omega$ : 随机试验所产生的所有样本点的集合

$\mathcal{F}$ : 集合内所有子集为元素的集合

$P(X)$ : 概率函数

**Axiom.** 非负性:

$$P(A) \geq 0, A \in \mathcal{F}.$$

**Axiom.** 规范性:

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

**Axiom.** 可列可加性: 对两两互斥的事件  $A_1, A_2, \dots$ ,

$$P\left(\sum_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i).$$

从三条公理得出的性质:

**Notation.** 1.  $P(\emptyset) = 0$

2. 有限可加性:

$$\sum_{i=1}^n P(A_i) = P\left(\sum_{i=1}^n A_i\right).$$

3.  $P(\bar{A}) = 1 - P(A)$

4.  $A \subset B \implies P(B - A) = P(B) - P(A)$

5.  $A \subset B \implies P(A) \leq P(B)$

6.  $P(A \cup B) = P(A) + P(B) - P(AB)$

**Notation.** 6.1.

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC).$$

6.2.

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n P(A_i A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n P(A_i A_j A_k) \\ \dots + (-1)^{n-1} P\left(\prod_{i=1}^n A_i\right).$$

**Example.** 从有号码  $1, 2, \dots, n$  的  $n$  个球中有放回地取  $m$  个球, 求取出的  $m$  个球中最大号码为  $k$  的概率

$$P\{k=1\} = \left(\frac{1}{n}\right)^m.$$

逐个列举计算较复杂, 记事件  $B_k$  为取出的  $m$  个球最大号码不超过  $k$ , 只需保证每次摸出的球都不超过  $k$  即可:

$$P(B_k) = \frac{k^m}{n^m}.$$

又有  $P(A_k) = P(B_k) - P(B_{k-1})$ , 且  $B_{k-1} \subset B_k$

所以:

$$P(A_k) = \frac{k^m}{n^m} - \frac{(k-1)^m}{n^m}.$$

**Example.** 匹配问题:  $n$  个学生各带有一个礼品, 随机分配礼品, 设第  $i$  个人抽到自己的礼品称为一个配对, 求至少有一个配对的概率

设  $A_i$  是第  $i$  个人抽到自己的礼品,  $A$  为目标事件, 则:

$$A = \bigcup_{i=1}^n A_i.$$

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

$$P(A_i A_j) = \frac{(n-2)!}{n!} = \frac{1}{P_n^2}.$$

$$P(A_i A_j A_k) = \frac{1}{P_n^3}.$$

.....

$$P\left(\prod_{i=1}^n A_i\right) = \frac{1}{n!}.$$

$$P(A) = P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n P(A_i A_j) + \dots$$

## 1.5 条件概率与乘法公式

**Definition.**

$$P(A) > 0, P(B|A) = \frac{P(AB)}{P(A)}.$$

即：在 **A** 发生的条件下，**B** 发生的概率

**Definition.** 乘法公式：

$$P(AB) = P(A)P(B|A) = P(B)P(A|B).$$

**Notation.**  $A, B$  独立：  $P(AB) = P(A)P(B)$

结合乘法公式：

$$P(B) = P(B|A).$$

$$P(A) = P(A|B).$$

## 1.6 全概率公式

**Corollary.** 事件  $A_1, A_2, \dots, A_n$  为完备事件组，事件  $B \subset \Omega = \bigcup_{i=1}^n A_i$ ，则：

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

**Notation.** 此时完备事件组的情况应该已知，通过完备事件组  $A$  的辅助可以求得较复杂事件  $B$  的概率

## 1.7 贝叶斯公式

**Corollary.**

$$P(A_k|B) = \frac{P(A_k B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_i)}.$$

贝叶斯公式被称为“逆概率公式/后验公式”，其中事件  $B$  更可能是事件的结果，将事件组  $A$  看作结果出现的原因，则贝叶斯公式是一个从“结果”推“原因”的可能性的公式

**Notation.** 对比一般公式：事件  $A$  导致  $B$ ，求  $B$  发生的概率

贝叶斯公式：事件  $A$  导致  $B$ ， $A$  中的一个事件  $A_i$  导致  $B$  发生的概率

**Axiom.** 条件概率的公理：

1. 非负性：  $P(A) \in [0, 1]$
2. 规范性：  $P(\Omega|A) = 1$
3. 可列可加性：

$$P\left(\sum_{i=1}^{\infty} B_i|A\right) = \sum_{i=1}^{\infty} P(B_i|A).$$

**Corollary.**

$$P(\bar{B}|A) = P(\Omega - B|A) = P(\Omega|A) - P(B|A).$$

**Corollary.**

$$\begin{aligned} P(B_1 \cup B_2) &= P(B_1) + P(B_2) - P(B_1 B_2). \\ \implies P(B_1 \cup B_2|A) &= P(B_1|A) + P(B_2|A) - P(B_1 B_2|A). \end{aligned}$$

**Corollary.** 乘法公式:

$$P(ABC) = P(A(BC)) = P(A)P(BC|A) = P(A)P(B|A)P(C|AB).$$

**Example.** 8 个红球 2 个白球, 求前三次结果是“红红白”的概率:

1. 不放回取 3 个 (和一次取三个球相同)

所有可能性:  $10 \times 9 \times 8$

目标事件:  $8 \times 7 \times 2$

或使用乘法公式: 设  $A_i$  为第  $i$  次取到红球, 目标事件可表示为  $A_1 A_2 \bar{A}_3$

概率:

$$P(A_1 A_2 \bar{A}_3) = P(A_1)P(A_2|A_1)P(\bar{A}_3|A_1 A_2) = \frac{8}{10} \times \frac{7}{9} \times \frac{2}{8} = \frac{7}{45}.$$

2. 每次取后放回, 并加入两个同色的球, 取 3 次 (不能使用古典概型)

概率:

$$P(A_1 A_2 \bar{A}_3) = \frac{8}{10} \times \frac{8}{12} \times \frac{2}{14} = \frac{8}{105}.$$

**Example.** 某疾病的发病率为 0.0004, 患病检测呈阳性的概率为 0.99, 误诊为阴性的概率为 0.01, 误诊为阳性的概率为 0.05, 不患病检测呈阴性概率为 0.95, 一个人检测呈阳性, 求其患病的概率

设阳性为  $A$ , 患病为  $B$

则:

$$P(A|B) = 0.99, P(A|\bar{B}) = 0.05, P(B) = 0.0004.$$

要求:  $P(B|A)$

使用贝叶斯公式:

$$\begin{aligned} P(B|A) &= \frac{P(AB)}{P(A)} = \frac{P(B)P(A|B)}{P(AB) + P(A\bar{B})}. \\ &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(A|\bar{B})P(\bar{B})} = 0.0079. \end{aligned}$$

## 1.8 独立性

**Definition.**  $A, B$  独立, 则:  $P(A|B) = P(A)$

**Notation.** 证明独立性:

1.  $P(A)P(B) = P(AB)$

**Notation.** 独立事件的特点:

1.  $A, B$  独立有:  $A, B$  所有的组合 (包含补集) 均独立
2.  $A, B$  独立的充要条件:  $P(A|B) = P(A)$  or  $P(B|A) = P(B)$
3.  $\emptyset$  与任何随机事件独立,  $\Omega$  与任何随机事件独立

对于三个事件相互独立:

$$\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(C)P(C) \\ P(ABC) = P(A)P(B)P(C) \end{cases}.$$

对比乘法公式:  $P(ABC) = P(A)P(B|A)P(C|AB)$

**Definition.** 相互独立:

有  $A_1, A_2, \dots, A_n$  事件组, 对  $\forall s \in [2, n]$  个事件  $A_{k_1}, A_{k_2}, \dots, A_{k_s}$  均有:

$$P\left(\prod_{n=1}^s A_{k_n}\right) = \prod_{n=1}^s P(A_{k_n}).$$

称事件  $A_1, A_2, \dots, A_n$  相互独立

**Definition.** 两两独立: 对事件  $A_1, A_2, \dots, A_n$ , 若任意两个事件独立, 则称为两两独立

**Notation.** 相互独立一定两两独立, 反之不一定

**Notation.** 相互独立事件组的性质:

1. 事件  $A_1, A_2, \dots, A_n$  相互独立, 将其中任意部分改为对立事件, 事件组仍为相互独立
2. 事件相互独立, 将事件组任意分为两组 (或多组), 对组内事件进行“并、交、差、补”操作后, 事件间依然相互独立

### 独立重复实验

**Definition.**  $E_1, E_2$  中一个试验的任何结果和另一个试验的任何结果相互独立, 则试验相互独立; 若  $n$  个独立试验相互独立且试验相同, 称  $E_1, E_2, \dots, E_n$  为  $n$  次独立重复实验, 或  $n$  重独立试验

**Example.** 扔硬币和掷骰子为独立试验, 其中扔硬币为伯努利试验 (只有两个结果)

**Definition.**  $n$  重独立试验  $E$  中, 每次试验都是伯努利试验 (可能结果只有两个), 称  $E$  为  $n$  重伯努利试验

1. 二项概率公式: 成功  $k$  次的概率记为  $P_n(k)$ , 假定前  $k$  次成功, 后  $n-k$  次失败, 则

$$P_i = p^k (1-p)^{n-k}.$$

指定事件  $A$  发生的位置有  $C_n^k$  种, 则:

$$P_n(k) = C_n^k p^k (1-p)^{n-k}.$$

称为二项概率公式

2. 几何概率公式: 首次成功恰好发生在第  $k$  次的概率记为  $G(k)$ , 设前  $k-1$  次失败, 则:

$$G(k) = q^{k-1} p.$$

可以验证:  $\sum G(k) = 1$

3. 负二项概率: 需要成功  $r$  次, 第  $r$  次成功恰好发生在第  $k$  次的概率记为  $G_r(k)$ , 设前  $k-1$  次试验有  $r-1$  次成功, 则:

$$G_r(k) = C_{k-1}^{r-1} p^r q^{k-r}.$$

同样有:  $\sum G_r(k) = 1$

## Lecture 3

## 2 一维随机变量及其分布

### 2.1 随机变量及其分布函数

随机变量

$$\text{随机变量} \begin{cases} \text{函数} \\ \text{连续} \end{cases}.$$

**Example.** 下一个进入教室的同学可能是男是女, 分别记为 1,2, 则有映射:

$$\{\text{男}, \text{女}\} \rightarrow \{1, 2\}.$$

将离散的结果映射为坐标轴上离散的数值, 所有的数值性的观测结果无需改变, 如: 下一个进入教室的同学身高为  $\omega$ , 则有映射:

$$X(\omega) = \omega.$$

**Definition.** 实值变量（无分布函数）使用小写字母，随机变量（有分布函数）使用大写字母  
对  $\forall x \in \mathbb{R}$ ,  $\{X \leq x\} = \{\omega | X(\omega) \leq x, \omega \in \Omega\} \in \mathcal{F}$ , 则  $X$  称为**概率空间的随机变量**

**Example.** 对于  $\{M, F\} \rightarrow \{1, 2\}$ , 取  $x = 1$ , 写出定义式:

$$\{X \leq 1\} = \{M\}.$$

同时由于  $x \in \mathbb{R}$ , 取  $x = 1.5$  时,  $\{X \leq 1.5\} = \{M\}$  取  $x = 4$  时,  $\{X \leq 4\} = \{M, F\}$

由于  $x \in \mathbb{R}$ , 则可以引入其他分布函数辅助, 继而引用微积分理论  
对于  $(\Omega, \mathcal{F}, P)$ :

$$P: \Omega \rightarrow [0, 1].$$

**Notation.**  $X$  具有随机性（样本点具有随机性），是定义在  $\Omega$  上的函数

$X$  是随机变量时  $\{a \leq X \leq b\}, a < b, a, b \in \mathbb{R}$  均为随机事件

$X$  是随机变量,  $g(x)$  是非单点的实值函数, 则  $Y = g(X)$  也是随机变量:

$$Y(\omega) = g(X(\omega)).$$

**Example.** 对灯泡做寿命试验, 用  $X$  表示测得灯泡的寿命, 样本空间  $\Omega = [0, +\infty)$ , 则:

$A =$  “测得灯泡寿命大于 500 h”  $= \{X > 500\}$

$B =$  “测得灯泡寿命小于 5000 h”  $= \{X \leq 5000\}$

## 分布函数

**Definition.** 分布函数: 记

$$F(x) = P\{X \leq x\}, x \in \mathbb{R}.$$

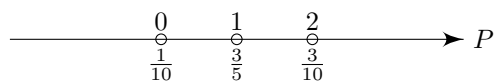
为  $X$  的分布函数

**Example.** 3 白 2 黑, 不放回取三次球, 求取到的黑球个数  $X$  的分布函数

$X$  可以取到: 0, 1, 2

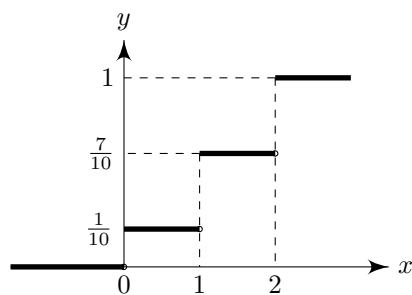
$$P\{X = 0\} = \frac{C_3^3}{C_5^3}, P\{X = 1\} = \frac{C_2^1 C_3^2}{C_5^3}, P\{X = 2\} = \frac{C_2^2 C_3^1}{C_5^3}.$$

概率在坐标轴上体现:



$$F(x) = P\{X \leq x\} = \begin{cases} 0, & x < 0 \\ \frac{1}{10}, & x \in [0, 1) \\ \frac{1}{10} + \frac{3}{5} = \frac{7}{10}, & x \in [1, 2) \\ 1, & x \geq 2 \end{cases}.$$

图像:



**Notation.** 分布函数的特性:

1. 非负性:  $P \in [0, 1]$
2. 单调不减性
3. 右连续性:

$$F(x) = \lim_{t \rightarrow x+0^+} F(t).$$

3.1 不满足左连续, 例:  $P(0) - P(0^-) \neq 0$

4. 规范性:

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1.$$

关于  $X$  的事件都可以使用分布函数表示:

$$\begin{cases} P\{X = a\} = \lim_{\varepsilon \rightarrow 0^+} P\{a - \varepsilon < X \leq a\} = F(a) - F(a - 0^+) \\ P\{a \leq X < b\} = F(b - 0^+) - F(a - 0^+) \\ \dots\dots \end{cases}.$$

**Example.** 在  $[a, b]$  内随机取一个数  $X$ , 求  $X$  的分布函数

关键区域:  $x \in [a, b]$ ,

$$\{X \leq x\} = \{a \leq X \leq x\}.$$

$$F(x) = P\{X \leq x\} = \frac{x - a}{b - a}.$$

作业: 预习第 2,3 节

## Lecture 4

$$\text{随机变量} \begin{cases} \text{离散型} \begin{cases} \text{有限型} \\ \text{无限型} \end{cases} \\ \text{连续型} \end{cases}.$$



**Notation.** 回忆: 分布函数有以下特征:

1. 非负性
2. 规范性
3. 右连续性
4.  $\forall x < y \in \mathbb{R}, F(x) \leq F(y)$

计算随机变量的概率可以用分布函数表达:

$$P\{a \leq X \leq b\} = F(b) - F(a).$$

## 2.2 常见分布律

1. 退化分布

2. 两点分布

- 2.1.  $0 \sim 1$  分布 ( $X \sim B(1, p)$ ):

$$P\{X = k\} = p^k (1-p)^{1-k}, k = 0, 1.$$

3. 二项分布 ( $X \sim B(n, p)$ ):

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n, p \in (0, 1).$$

4. 几何分布 ( $X \sim G(p)$ ):

$$P\{X = k\} = p(1-p)^{k-1}, k = 1, 2, 3, \dots, p \in (0, 1).$$

5. 泊松分布 ( $X \sim P(\lambda)$ ): 用于描述稀有事件的发生

$$P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots, \lambda > 0.$$

**Notation.** 由

$$e^x = \sum_{i=0}^{+\infty} \frac{x^i}{i!}.$$

可得:

$$\sum_{k=0}^{+\infty} P\{X = k\} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

**Notation.** 分布律的基本性质:

1. 非负性:  $p_i \geq 0$
2. 正则性:  $\sum_{i=1}^{+\infty} p_i = 1$ , 即每一个点的概率都应该知道

**Example.** 保险问题

若一年中某类保险投保人死亡的概率为 0.005, 现有 10000 人参加保险, 求未来一年中:

1. 40 人死亡的概率

设  $X$  为未来一年中死亡的人数, 有  $X \sim B(10000, 0.005)$ , 计算:

$$P\{X = 40\} = C_{10000}^{40} 0.005^{40} \cdot 0.995^{9960} \approx 2.143 \times 10^{-2}.$$

直接计算较为复杂, 可以使用近似计算

有两种近似计算方法: 泊松定理、中心极限定理

**Notation.** 泊松定理: 二项分布有时可以转化为泊松分布:

如果  $\lim_{n \rightarrow +\infty} np_n = \lambda > 0$  (极小但不为 0), 则:

$$\lim_{n \rightarrow +\infty} C_n^k p_n^k (1 - p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots$$

前提:  $n$  大  $p$  小

将保险问题转换为泊松分布:

$$\lambda = np = 50.$$

$$P\{X = 40\} = \frac{50^{40}}{40!} e^{-50} \approx 0.02.$$

2. 死亡人数不超过 70 的概率

$$\begin{aligned} P\{X \leq 70\} &= \sum_{k=0}^{70} C_{10000}^k 0.005^k \cdot 0.995^{(10000-k)} \\ &= \sum_{k=0}^{70} \frac{50^k}{k!} e^{-50} (\lambda = np = 50). \end{aligned}$$

**Notation.** 几何分布具有无记忆性: 当前试验对过去的试验无任何影响, 即:

$$P\{X = k + 1 | X > k\} = P\{X = 1\}.$$

可以使用条件概率证明:

$$P\{X = k + 1 | X > k\} = \frac{P\{X = k + 1, X > k\}}{P\{X > k\}}.$$

由于:

$$P\{X > k\} = \sum_{j=k}^{+\infty} p(1-p)^j = p \sum_{j=k}^{+\infty} (1-p)^j = p(1-p)^k \cdot \frac{1}{1-(1-p)} = (1-p)^k.$$

## 2.3 连续型随机变量

**Definition.**

$$F(x) = \int_{-\infty}^x f(x) dx, x \in \mathbb{R}.$$

则  $X$  为连续型随机变量,  $f(x)$  称为  $X$  的密度函数

连续型随机变量的性质:

1. 非负性:  $f(x) \geq 0, x \in \mathbb{R}$
2. 规范性:

$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

- 3.

$$P\{a < X \leq b\} = \int_a^b f(x) dx, a < b.$$

4.  $F$  连续

5.  $F'(x) = f(x)$

**Notation.** 由于连续性随机变量的分布函数  $F$  处处连续, 所以  $\forall x \in \mathbb{R}$ , 有  $P\{X = x\} = F(x) - F(x-0) = 0$ , 即: 概率为 0 的事件不一定是不可可能事件

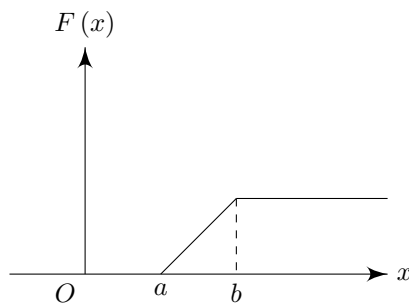
**Example.**

常见的连续型密度函数:

1. 均匀分布 ( $X \sim U[a, b]$ ):

$$f(x) = \begin{cases} \frac{1}{b-a}, x \in [a, b] \\ 0, x \notin [a, b] \end{cases}.$$

对应的分布函数图像:



2. 指数分布 ( $X \sim \Gamma(1, \lambda)$ ):

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x > 0 \\ 0, x \leq 0 \end{cases}.$$

**Notation.** 指数分布大多数与等待时间有关

指数分布的充分必要条件为

$$\forall s, t \geq 0, P\{X > s + t | X > s\} = P\{X > t\}.$$

即指数分布有无记忆性/无后效型 (指数分布的特点)

3. 正态分布 ( $X \sim N(\mu, \sigma^2)$ ):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

**Notation.**  $\sigma^2$ : 方差,  $\mu$ : 数学期待

3.1. 标准正态分布 ( $X \sim N(0, 1)$ ):

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

## Lecture 5

**Notation.** 马尔可夫分布也具有无记忆性

回忆: 正态分布 (高斯分布)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

标准正态分布

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

## 2.4 标准化

**Definition.** 标准化将随机变量转化为另一组随机变量

对于  $X \sim N(\mu, \sigma^2)$ , 移除前者的中心 (将中心变为 0), 除以标准差, 即得到符合标准正态分布的  $Y \sim N(0, 1)$

1. 去中心化

2. 除以标准差

即:

$$Y = \frac{X - E(X)}{\sqrt{D(X)}}.$$

标准化后的随机变量期望为 0, 方差为 1

$$P\{|X - \mu| < k\sigma\} = P\left\{\left|\frac{x - \mu}{\sigma}\right| < k\right\} = P\{Y < k\}.$$

此时  $Y \sim N(0, 1)$  , 即符合标准正态分布

对于原来的  $3\sigma$  原则, 转化为  $P\{-3 < Y < 3\}$

$$\begin{aligned} P\{-3 < Y < 3\} &= F_Y(3) - F_Y(-3). \\ &= \phi(3) - \phi(-3). \end{aligned}$$

**Notation.**

$$\phi(-x) = 1 - \phi(x).$$

$$= 2\phi(3) - 1.$$

**Example.** 设人的高度符合正态分布  $X \sim N(170, 49)$ , 问在公共设施处的门需要设计多高才能使至少 90% 的人通过

求门高  $H$  使得  $P\{X \leq H\} \geq 0.9$

$$P\{X \leq H\} \implies P\left\{\frac{X - 170}{49} \leq \frac{H - 170}{49}\right\}.$$

即:

$$\begin{aligned} \phi\left(\frac{H - 170}{49}\right) &\geq 0.9. \\ \frac{H - 170}{49} &\geq 1.28 \implies H \geq 1.80. \end{aligned}$$

## 2.5 随机变量函数的分布

**Example.**

$$\begin{aligned} Y &= \frac{X - E(X)}{\sqrt{D(X)}}. \\ Y = g(X) &\sim F_Y(Y). \end{aligned}$$

求解  $F_Y(Y)$

**Example.**  $D \sim U[a, b]$  且

$$S = \pi \left(\frac{b}{2}\right)^2 = \frac{\pi \rho^2}{4}.$$

1.  $X$  离散:  $Y$  一般是离散的
2.  $X$  连续:  $Y$  可能连续, 可能分段连续 (离散)

表 1: Y Func

0.15	0.1	0.1	0.2	0.3	0.15
4	1	0	1	4	9

**Example.**  $X$  的分布律:

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 & 3 \\ 0.15 & 0.1 & 0.1 & 0.2 & 0.3 & 0.15 \end{pmatrix}.$$

求  $Y = X^2$  的分布律

列举  $Y$  的分布律: 合并后:

$$\begin{pmatrix} 0 & 1 & 4 & 9 \\ 0.1 & 0.3 & 0.45 & 0.15 \end{pmatrix}.$$

**Example.** 分析法:  $X \sim G(0.5)$  (几何分布), 求  $Y = \sin\left(\frac{\pi}{2}X\right)$  的分布律

易得:  $Y$  可以取得: 0, 1, -1

$$Y = \sin\left(\frac{\pi}{2}X\right) = \begin{cases} -1, X = 4n - 1 \\ 0, X = 4n \text{ 且 } X = 4n - 2 \\ 1, X = 4n - 3 \end{cases}.$$

$$P\{Y = -1\} = \sum_{n=1}^{+\infty} P\{X = 4n - 1\} = \sum_{n=1}^{+\infty} 0.5 \times 0.5^{4n-1-1}.$$

同理:

$$P\{Y = 0\} = \sum_{n=1}^{+\infty} 0.5 \times 0.5^{2n-1}.$$

$$P\{Y = 1\} = \sum_{n=1}^{+\infty} 0.5 \times 0.5^{4n-4}.$$

求得  $Y$  的分布律:

$$\begin{pmatrix} -1 & 0 & 1 \\ \frac{2}{15} & \frac{1}{3} & \frac{8}{15} \end{pmatrix}.$$

## Lecture 6

**Corollary.**  $X$  是连续性随机变量, 密度函数为  $f_X(x)$

随机变量  $Y = g(X)$ , 且  $\exists D, P\{Y \in D\} = 1$ ,  $g(x)$  存在反函数  $h(y)$  且严格单调可导, 则:

$$f_Y(y) = \begin{cases} |h'(y)| f_X(h(y)), y \in D \\ 0, \text{Others} \end{cases}.$$

**Notation.** 指数分布  $X \sim I(1, \lambda)$  的数学期望  $E(X) = \frac{1}{\lambda}$

### 3 多维随机变量函数及其分布

在实际问题中, 试验结果有时需要使用两个或两个以上的随机变量 (random value, r.v.) 来描述

**Example.** 天气预报: 温度、湿度、风力、降水等

#### 3.1 二维随机变量及其分布

**Definition.** 设  $\Omega$  为随机试验的样本空间, 则

$$\forall \omega \in \Omega \xrightarrow{\text{某种变化}} \exists (X(\omega), Y(\omega)) \in \mathbb{R}^2.$$

或:

$$\{X \leq x, Y \leq y\} = \{\omega | X(\omega) \leq x, Y(\omega) \leq y, \omega \in \Omega\} \in \mathcal{F}.$$

称  $(X, Y)$  为概率空间  $(\Omega, \mathcal{F}, P)$  上的二维随机变量

**Notation.**  $\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$

性质:

1.  $F(x, y) \in [0, 1]$
2. 关于每个变量单调不减, 即固定  $x$ , 对  $\forall y_1 < y_2$ ,

$$F(x, y_1) \leq F(x, y_2).$$

3. 对每个变量右连续, 即:

$$F(x_0, y_0) = F(x_0 + 0^+, y_0) = F(x_0, y_0 + 0^+).$$

4. 对  $\forall a < b, c < d$ , 有:

$$F(b, d) - F(b, c) - F(a, d) + F(a, c) \geq 0.$$

即: 在任意地方框一个矩形, 内部区域的概率必须大于等于 0

**Example.** 性质 4 例题: 设

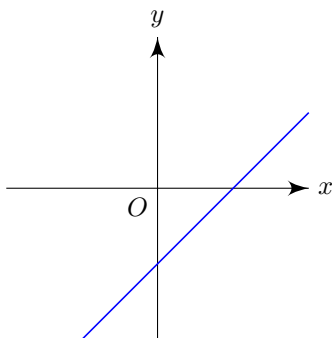
$$F(x, y) = \begin{cases} 0, & x + y < 1 \\ 1, & x + y \geq 1 \end{cases}.$$

讨论  $F(x, y)$  能否成为二维随机变量的分布函数

**Notation.**

$$P\{X > a, Y > c\} \neq 1 - F(a, c).$$

图 2: 分布函数图像



### 3.2 二维随机变量的边缘分布函数

边缘分布: 降一维

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\{X \leq x, Y < +\infty\} \\ &= F(x, +\infty) \end{aligned}$$

**Example.** 设随机变量  $(X, Y)$  的联合分布函数为

$$F(x, y) = A \left( B + \arctan \frac{x}{2} \right) \left( C + \arctan \frac{y}{2} \right), x, y \in (-\infty, +\infty).$$

求  $A, B, C$

解:  $\arctan x$  的性质:

$$\lim_{x \rightarrow \pm\infty} \arctan x = \pm \frac{\pi}{2}.$$

则

$$F(+\infty, +\infty) = A \left( B + \frac{\pi}{2} \right) \left( C + \frac{\pi}{2} \right) = 1.$$

$$F(-\infty, +\infty) = A \left( B - \frac{\pi}{2} \right) \left( C + \frac{\pi}{2} \right) = 0.$$

$$F(-\infty, -\infty) = A \left( B - \frac{\pi}{2} \right) \left( C - \frac{\pi}{2} \right) = 1.$$

联立解出  $A, B, C$

### 3.3 联合分布律

二维离散随机变量的联合分布函数

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}.$$



表 2: 联合分布律

$X \setminus Y$	$b_1$	$b_2$	$\dots$	$b_j$	$p_{a \cdot}$
$a_1$	$p_{11}$	$p_{12}$	$\dots$	$p_{1j}$	$\sum_{n=1}^j p_{n1}$
$a_2$	$p_{21}$	$p_{22}$	$\dots$	$p_{2j}$	$\sum_{n=2}^j p_{n2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$a_i$	$p_{i1}$	$p_{i2}$	$\dots$	$p_{ij}$	$\sum_{n=i}^j p_{ni}$
$p_{b \cdot}$	$\sum_{m=1}^i p_{im}$	$\sum_{m=2}^i p_{im}$	$\dots$	$\sum_{m=1}^i p_{im}$	1

如何求  $p_{ij}$  :

1. 古典概型
2. 乘法公式:

$$p_{ij} = P\{X = x_i\} P\{Y = y_j | X = x_i\} = P\{X = x_i, Y = y_j\}.$$

### 3.4 二维连续性随机变量及其概率特性

**Definition.** 若

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du.$$

则称  $f(x, y)$  为二维随机变量  $(X, Y)$  的联合密度函数, 称  $(X, Y)$  为二维连续型随机变量

**Notation.** 联合密度与联合分布函数的性质:

1.  $f(x, y) \geq 0$
- 2.

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dy dx = 1 = F(-\infty, +\infty).$$

3. 对每个边缘连续, 在  $f(x, y)$  的连续点处:

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y).$$

从而有:  $P(x < X \leq x + \Delta x, y < Y \leq y + \Delta y) \approx f(x, y) \Delta x \Delta y$

## Lecture 7

### 两个随机变量的独立性

证明. 当  $A, B$  独立时:  $P(AB) = P(A)P(B)$

$$\forall i, j: P\{X = x_i, Y = y_j\} = P\{X = x_i\} P\{Y = y_j\}.$$

□

**Notation.** 离散随机变量独立的情况下:  $P_{ij} = P_{i \cdot} \cdot P_{\cdot j}$

如何证明  $X, Y$  独立:

$$F(x, y) = P\{X \leq x, Y \leq y\} = P\{X \leq x\} P\{Y \leq y\} = F_X(X) \cdot F_Y(y).$$

由联合分布律得边缘分布律:

$$F_X(x) = \lim_{y \rightarrow +\infty} F(x, y) = \lim_{y_0 \rightarrow +\infty} P\{X \leq x, Y \leq y_0\}.$$

**Notation.** 连续性随机变量的区间  $D$  概率:

$$F(x, y) = P\{X \leq x, Y \leq y\} = \iint_{u=x, v=y} f(u, v) dx dy.$$

**概率函数**

**Notation.** 二维均匀分布:

$$f(x, y) = \begin{cases} \frac{1}{S(D)}, & (x, y) \in D \\ 0, & (x, y) \notin D \end{cases}.$$

**Example.**

$$f(x, y) = \begin{cases} Axy, & x \in (0, 1), y \in (0, 1) \\ 0, & \text{Others} \end{cases}.$$

1. 求  $A$

$$\iint_{x \in [0, 1], y \in [0, 1]} Axy dx dy = 1.$$

$$A = 8$$

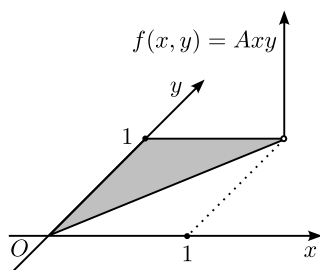
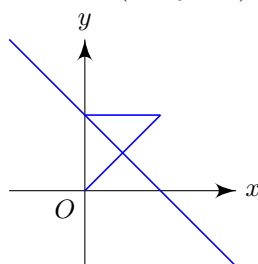


图 3: 函数积分域

2.  $P\{X + Y \geq 1\}$

$$\begin{aligned}
 P\{X + Y \geq 1\} &= \iint_{x+y \geq 1} f(x, y) \, dx dy \\
 &= \iint_{x+y \geq 1} 8xy \, dx dy \\
 &= \int_{0.5}^1 dy \int_{1-y}^y 8xy \, dx \\
 &= \frac{5}{6}.
 \end{aligned}$$

图 4:  $P(x + y \geq 1)$



3.  $X, Y$  的分布函数

3.1  $x \in (-\infty, 0), y \in (-\infty, 0)$

3.2  $x \in [0, 1), y \in [0, x)$

3.3  $x \in [0, 1), y \in [x, 1)$

3.4  $x \in [0, 1), y \in [1, +\infty)$

3.5  $x \in [1, +\infty), y \in [0, 1)$

3.6  $x \in [1, +\infty), y \in [1, +\infty)$  分段对  $Axy$  积分:

$$F(x, y) = \iint_{x \leq u, y \leq v} f(u, v) \, dx dy$$

### 3.5 多维随机变量及分布

**Definition.** 二维推广至多维:

$$F(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}.$$

称  $F$  为  $n$  维随机变量的联合分布函数

**Definition.** 多维联合分布律:

$$P\{X_1 = a_{1k_1}, X_2 = a_{2k_2}, \dots, X_n = a_{nk_n}\}.$$

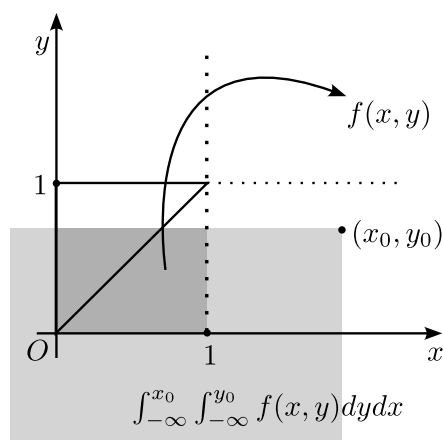


图 5: 积分区域

联合分布函数和联合分布律的关系:

$$F(x_1, x_2, \dots, x_n) = \sum_{a_{1k_1} \leq x_1} \sum_{a_{2k_2} \leq x_2} \cdots \sum_{a_{nk_n} \leq x_n} P\{X_1 = a_{1k_1}, X_2 = a_{2k_2}, \dots, X_n = a_{nk_n}\}.$$

**Notation.** 二项分布推广多项分布:

$A_1, A_2, \dots, A_r$  是  $E$  的完备事件组,  $P(A_i) = p_i, i = 1, 2, \dots, r$ , 对  $E$  进行  $n$  次独立重复试验,  $X_i$  表示  $A_i$  发生的次数, 则:

$$P\{X_1 = k_1, X_2 = k_2, \dots, X_r = k_r\} = \frac{n!}{k_1! k_2! \cdots k_r!} \prod_{i=1}^r p_i^{k_i}.$$

其中  $k_i \geq 0, \sum_{i=1}^r k_i = n$ , 当  $n = 2$  时为二项分布

### 3.5.1 多维随机变量的独立性

**Definition.**

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i).$$

称随机变量相互独立

等价于:

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i).$$

## Lecture 8

10.15

## 3.5.2 条件分布

**Notation.** 离散型条件分布:

$$P\{X = a_i | Y = b_j\} = \frac{P\{X = a_i, Y = b_j\}}{P\{Y = b_j\}}.$$

$$P\{Y = b_j\} = \sum_{i=1}^{+\infty} P\{Y = b_j, X = a_i\}.$$

性质:

1.  $P\{X = a_i | Y = b_j\} \geq 0$
2.  $\sum_{i=1}^{+\infty} P = 1$

**Notation.** 连续型条件分布:

$$P\{X = a | Y = b\} = 0.$$

(无穷多个点)

通过微元法:

$$\begin{aligned} P\{X \leq x | Y = y\} &= \lim_{\varepsilon \rightarrow 0^+} P\{X \leq x | y - \varepsilon < Y \leq y\} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{P\{X \leq x, Y \in (y - \varepsilon, y]\}}{P\{Y \in (y - \varepsilon, y]\}} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{F(x, y) - F(x, y - \varepsilon)}{F_Y(y) - F_Y(y - \varepsilon)} \\ &= \frac{\frac{\partial F(x, y)}{\partial y}}{\frac{dF_Y(y)}{dy}} \\ &= \frac{\int_{-\infty}^x f(u, y) du}{f_Y(y)} \\ &= \int_{-\infty}^x \frac{f(u, y)}{f_Y(y)} du. \end{aligned}$$

## 3.6 二维随机变量函数的分布

$$\begin{cases} aX + bY + c \\ \max\{X, Y\} \\ \min\{X, Y\} \end{cases}.$$

重点公式: 3.4.5 ~ 3.4.8

假设随机变量  $Z = aX + bY + c$ , 有分布函数  $(X, Y) \rightarrow F(x, y)$

$$\begin{aligned}
F_Z(z) &= P\{Z \leq z\} \\
&= P\{aX + bY + c \leq z\} \\
&= \begin{cases} \sum_{ax_i + by_j + c \leq z} P\{X = x_i, Y = y_j\}, & (X, Y) \text{ 离散} \\ \iint_{ax+by+c \leq z} P\{X = x, Y = y\} dx dy, & (X, Y) \text{ 连续} \end{cases}.
\end{aligned}$$

**Notation.** 二项分布可加性: 有两个相互独立的试验  $X \sim B(m, p), Y \sim B(n, p)$ , 相当于一个试验  $Z \sim B(m+n, p)$

泊松分布可加性: 相互独立的两个随机变量  $X \sim P(\lambda_1), Y \sim P(\lambda_2)$ , 相当于一个分布  $Z \sim P(\lambda_1 + \lambda_2)$

**Notation.** 极值公式:

两个随机变量 (连续)  $X, Y$  相互独立, 求  $Z_1 = \max(X, Y), Z_2 = \min(X, Y)$  的分布函数和密度函数

1.  $F_{Z_1}(z)$ : 最大的不超过  $z$  等价于每一个都不超过  $z$

$$\begin{aligned}
F_{Z_1}(z) &= P\{X \leq z, Y \leq z\} \\
&= P\{X \leq z\} \cdot P\{Y \leq z\} \\
&= F_X(z) \cdot F_Y(z).
\end{aligned}$$

2.  $F_{Z_2}(z)$ : 最小的不超过  $z$  不等价于每一个都不超过  $z$ , 但最小的超过  $z$  等价与每一个都超过了  $z$

$$\begin{aligned}
F_{Z_2}(z) &= 1 - P\{X > z, Y > z\} \\
&= 1 - [1 - F_X(z)] \cdot [1 - F_Y(z)].
\end{aligned}$$

**Notation.** 独立同分布: 变量相互独立且分布律相同

对极值公式扩展:  $(X_1, X_2, \dots, X_n)$  独立同分布:

$$F_{Z_1}(z) = \prod_{i=1}^n F_{X_i}(z).$$

由于同分布, 因此  $F_{X_i}(z) = F_X(z)$

$$F_{Z_1}(z) = (F_X(z))^n.$$

$$f_{Z_1}(z) = (F_{Z_1}(z))' = n(F_X(z))^{n-1}.$$

**Notation.** 3.4.5:

$$\begin{aligned}
 F_Z(z) &= P\{X + Y \leq z\} \\
 &= \iint_D f(x, y) \, dx \, dy \\
 &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x, y) \, dy \\
 &= \int_{-\infty}^z dt \int_{-\infty}^{+\infty} f(x, t-x) \, dx \\
 f_Z(z) &= F'_Z(z) \\
 &= \int_{-\infty}^{+\infty} f(x, z-x) \, dx.
 \end{aligned}$$

同理:

$$f_Z(z) = \int_{-\infty}^{+\infty} f(z-y, y) \, dy.$$

称以上两个公式为卷积公式:  $f_Z = f_X * f_Y$

**Notation.** 正态分布的可加性:

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ , 且  $X, Y$  相互独立, 则:

$$X + Y = Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

多元:

$$Z = \sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$

### 3.7 二元正态分布

$$\begin{aligned}
 f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right. \\
 &\times \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\Big\}.
 \end{aligned}$$

## Lecture 9

10.17

**Example.**  $X \sim U[-1, 1]$ , 对应的密度函数:

$$f_X(x) = \begin{cases} 0, & x < -1 \text{ or } x > 1 \\ \frac{1}{2}, & x \in [-1, 1] \end{cases}.$$

同理  $Y \sim U[-1, 1]$ , 求  $Z = |X - Y|$  的分布函数

解:  $Z$  的取值:  $[0, 2]$

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} \\ &= P\{|X - Y| \leq z\} \\ &= \begin{cases} 0, & z < 0 \\ 1, & z \geq 2 \\ P\{-z \leq X - Y \leq z\}, & z \in [0, 2) \end{cases}. \end{aligned}$$

其中

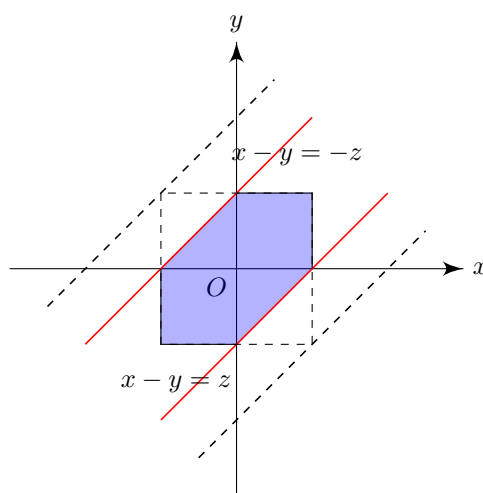
$$P\{-z \leq X - Y \leq z\} = \iint_{-z \leq x-y \leq z} f(x, y) dx dy.$$

通过  $f_X$  和  $f_Y$  求联合密度函数:  $X, Y$  独立, 即  $f(x, y) = f_X(x) f_Y(y)$

$$f(x, y) = \begin{cases} 0, & x, y < -1 \text{ or } x, y > 1 \\ \frac{1}{4}, & x, y \in [-1, 1] \end{cases}.$$

$$P\{-z \leq X - Y \leq z\} = \iint_{-z \leq x-y \leq z, x, y \in [-1, 1]} \frac{1}{4} dx dy.$$

画图确认积分区域:



**Notation.** i.i.d. : 独立同分布

**Notation.** 伽马分布  $\Gamma(\alpha, \beta)$  的密度函数:

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

**Notation.** 重点题目: 3.6.2



## 4 数字特征

- 数学期望:  $E(X)$
- 方差:  $D(X)$  or  $\text{Var}(X)$
- 协方差:  $\text{cov}(X, Y)$
- 相关系数:  $\rho(X, Y)$
- 矩:  $E(X)^k$  and  $E(X - EX)^k$

$$D(X) = E(X - EX)^2.$$

$$\text{cov}(X, Y) = E((X - EX)(Y - EY)).$$

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}}.$$

**Notation.**  $|\rho_{X,Y}| \in [0, 1]$

$\rho$  越大越线性相关,  $\rho > 0.8$  时基本可以确定为线性相关

**Notation.** 矩 (moment) 是最一般的概念

矩分为两大类:  $k$  阶原点矩和  $k$  阶中心矩

原点矩:  $E(X)^k$

中心矩:  $E(X - EX)^k$

$k+l$  阶混合中心矩:  $E((X - EX)^k (Y - EY)^l)$

**Example.** 数学期望为一阶原点矩

方差为一阶中心矩

协方差为二阶混合中心矩

可以写出无穷阶的中心矩等同于通过泰勒原理得出分布函数

本章重点: 如何计算任意随机变量有关函数的数学期望

唯一计算公式: 4.1.5 和 4.1.6

### 4.1 数学期望

**Definition.** 离散型随机变量  $X$  的分布律:  $P\{X = x_i\} = p_i, i = 1, 2, \dots$

若级数  $\sum_{i=1}^{+\infty} x_i p_i$  **绝对收敛** ( $\sum_{i=1}^{+\infty} |x_i| p_i < +\infty$ ), 则  $X$  的数学期望**存在** ( $x_i$  为取值,  $p_i$  为权重,  $p_i \geq 0$ )

$$E(X) = EX = \sum_{i=1}^{+\infty} x_i P\{X = x_i\} = \sum_{i=1}^{+\infty} x_i p_i.$$

**Rule.** 当一个随机变量的密度函数与分布律已知:  $X \rightarrow f(x), P\{X = x_i\} = p_i$

即可以求关于  $X$  函数的数学期望 (公式 4.1.5):

$$E(g(X)) = \begin{cases} \sum_{i=1}^{+\infty} g(x_i) P\{X = x_i\}, & X \text{ 离散} \\ \int_{-\infty}^{+\infty} g(x) f(x) dx, & X \text{ 连续} \end{cases}.$$

**Rule.** 扩展至二阶:  $(X, Y) \rightarrow P\{X = x_i, Y = y_j\}, f(x, y)$

关于  $(X, Y)$  的函数的数学期望 (公式 4.1.6):

$$E(g(X, Y)) = \begin{cases} \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} g(x_i, y_j) P\{X = x_i, Y = y_j\}, & \text{离散} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy, & \text{连续} \end{cases}$$

**Notation.** 柯西分布:

$$f(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}.$$

常见分布数学期望:

**Notation.** 伯努利分布  $X \sim B(n, p): EX = np$

泊松分布  $X \sim P(\lambda) (\lambda > 0): EX = \lambda$

柯西分布:  $EX$  不存在 (柯西分布不绝对收敛)

**Notation.** 柯西活了 68 岁, 21 岁成名 (导师拉格朗日), 27 岁当选法国科学院院士

## Lecture 10

10.22

**Example.**  $(X, Y) \sim N_2(0, 1), \phi(x, y) = \frac{1}{2\pi} e^{-x^2+y^2/2}$ , 令  $Z = \sqrt{X^2 + Y^2}$ , 求  $E(Z)$

解: 由定理:

**Rule.**

$$E(g(X, Y)) = \begin{cases} \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} g(x_i, y_j) P\{X = x_i, Y = y_j\}, & \text{离散} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy, & \text{连续} \end{cases}$$

可得数学期望:

$$E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Z \cdot f(x, y) dx dy.$$

### 4.2 数学期望的性质

- 线性可加性
- 独立性

◦  $E(aX + bY + c) = E(aX + bY) + c$ : 常数的数学期望为其本身

**Notation.** 什么是数学期望: 一个随机变量的中心

方差: 去中心化的随机变量

常数的中心为其本身

◦  $E(aX + bY) = E(aX) + E(bY) = aE(X) + bE(Y)$ : 线性性

证明. 已知:

$$\int_{-\infty}^{+\infty} f(x, y) dy = F_X(x).$$

$$\int_{-\infty}^{+\infty} xf(x, y) dx = E(X).$$

$$\begin{aligned} E(aX + bY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (ax + by) f(x, y) dx dy \\ &= a \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y) dx dy + b \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yf(x, y) dx dy \\ &= aE(X) + bE(Y). \end{aligned}$$

□

**Example.**  $E(X) \pm E(Y) = E(X \pm Y)$

◦ 对于独立的随机变量:  $E(XY) = E(X) \cdot E(Y)$

证明. 二重积分转换为二次积分:

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_X(x) f_Y(y) dx dy \\ &= \left( \int_{-\infty}^{+\infty} xf_X(x) dx \right) \left( \int_{-\infty}^{+\infty} yf_Y(y) dy \right) \\ &= E(X) \cdot E(Y). \end{aligned}$$

□

**Notation.**

$$\begin{aligned} \text{cov}(X, Y) &= E(XY) - E(X)E(Y) = 0 \\ \implies \rho_{X,Y} &= \frac{\text{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} = 0 \\ \implies X, Y &\text{无相关 (独立)}. \end{aligned}$$

**Notation.** 线性可加性:

$$E\left(\sum_{i=1}^n a_i X_i + c\right) = \sum_{i=1}^n a_i E(X_i) + c.$$

将  $n$  重积分转换为一重积分

### 4.3 方差的性质

$$D(X) = E(X - EX)^2.$$

**Notation.** 方差是描述数据偏离中心的程度值

◦ 常数的方差等于 0:  $D(c) = 0$

**Notation.** 正态分布的方差  $\sigma$  不大:  $3\sigma$  准则保证数据方差在可控范围内

◦  $D(aX + b) = D(aX) = a^2 D(X)$ : 离散程度与整体移动无关

证明.

$$\begin{aligned} D(aX + b) &= E(aX + b - E(aX + b))^2 \\ &= E(aX + b - aE(X) - b)^2 \\ &= E(aX - aE(X))^2 = D(aX) \\ &= a(X - E(X)) \cdot aE(X - E(X)) \\ &= a^2 E(X - E(X))^2 \\ &= a^2 D(X). \end{aligned}$$

□

◦  $D(X \pm Y) = D(X) + D(Y) \pm 2E((X - EX) \cdot (Y - EY))$

证明.

$$\begin{aligned} D(X - Y) &= E(X - Y - E(X - Y))^2 \\ &= E(X - Y - (EX - EY))^2 \\ &= E((X - EX) - (Y - EY))^2 \\ &= E\left((X - EX)^2 - 2(X - EX)(Y - EY) + (Y - EY)^2\right) \\ &= E(X - EX)^2 - 2E(X - EX)(Y - EY) + E(Y - EY)^2 \\ &= D(X) + D(Y) - 2\text{cov}(X, Y). \end{aligned}$$

$$\begin{aligned} \text{cov}(X, Y) &= E(X - EX)(Y - EY) \\ &= E(XY - X \cdot EY - Y \cdot EX + EX \cdot EY) \\ &= E(XY) - E(X \cdot EY) - E(Y \cdot EX) + E(EX \cdot EY) \\ &= E(XY) - EY \cdot E(X) - EX \cdot E(Y) + EX \cdot EY \\ &= E(XY) - E(X) \cdot E(Y). \end{aligned}$$

当  $X, Y$  独立时:  $\text{cov}(X, Y) = 0$ , 即  $D(X - Y) = D(X) + D(Y)$ , 加法同理  $\square$

**Notation.** 当  $X = Y$  时:

$$\begin{aligned}\text{cov}(X, Y) &= \text{cov}(X, X) \\ &= E(X - EX)(X - EX) \\ &= E(X - EX)^2 \\ &= D(X).\end{aligned}$$

即协方差退化为方差

**Notation.** 均方偏离函数:  $f(x) = E(X - x)^2 \geq D(X)$ , 当且仅当  $x = E(X)$  时  $f(X) = D(X)$

◦ 切比雪夫不等式 (概率论最基础的不等式)

$$P\{|X - EX| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}.$$

或:

$$P\{|X - EX| > \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon}.$$

证明时使用:

$$P\{(X - EX)^2 \leq \varepsilon^2\} \leq \frac{D(X)}{\varepsilon^2}.$$

证明.

$$\begin{aligned}P\{|X - EX| \geq \varepsilon\} &= \int_{|x - EX| \geq \varepsilon} f(x) dx \\ &\leq \int_{|x - EX| \geq \varepsilon} \frac{|x - EX|^2}{\varepsilon^2} f(x) dx \\ &\leq \int_{-\infty}^{+\infty} \frac{|x - EX|^2}{\varepsilon^2} f(x) dx \\ &= \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx \\ &= \frac{1}{\varepsilon^2} E(X - EX)^2 \\ &= \frac{D(X)}{\varepsilon^2}.\end{aligned}$$

$\square$

**Notation.** 切比雪夫不等式  $\implies$  马尔可夫不等式  $\implies$  协方差不等式  $\implies$  阶乘不等式  $\implies \dots$

$D(X) = 0$  的充要条件为  $P = 1$

## Lecture 11

10.24

*Review:***Notation.** 数学期望的性质:

1.  $E(c) = c$
2.  $E(cX) = cE(X)$
3.  $E(X + Y) = E(X) + E(Y)$
- 3.1  $E(E(Y)X) = E(Y)E(X)$
4.  $X, Y$  相互独立,  $E(XY) = E(X)E(Y)$

协方差:  $\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - E(X)E(Y)$ 若  $X, Y$  独立则  $\text{cov}(X, Y) = 0$ **Notation.** 方差的性质:

1.  $D(c) = 0$
2.  $D(cX) = c^2D(X)$
- 2.1.  $D(X) = E(X - EX)^2 = E(X^2) - E(X)^2$
3.  $X, Y$  相互独立,  $D(X + Y) = D(X) + D(Y)$

 $\text{cov}(X, Y) = E(X - EX)(Y - EY)$ 当  $X = Y$ ,  $\text{cov}(X, Y) = \text{cov}(X, X) = E(X - EX)^2 = D(X)$ 或:  $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^2) - E(X)^2$ **Example.**  $D(aX + bY + c) = D(aX + bY)$ 

$$\begin{aligned}
 D(aX + bY) &= E((aX + bY) - E(aX + bY))^2 \\
 &= E(a(X - EX) + b(Y - EY))^2 \\
 &= E\left(a^2(X - EX)^2 + 2ab(X - EX)(Y - EY) + b^2(Y - EY)^2\right) \\
 &= a^2D(X) + b^2D(Y) + 2ab\text{cov}(X, Y).
 \end{aligned}$$

◦ 切比雪夫不等式: 已知一个随机变量的方差可以估算出数学期望

**Question.** 一个随机变量  $X$  分布未知, 已知  $\mu = 18, \sigma = 2.5$ , 求  $P\{X \in (8, 28)\}$ 

解: 由切比雪夫不等式:

$$\begin{aligned}
 P\{X \in (8, 28)\} &= P\{X - 18 \in (-10, 10)\} \\
 &= P\{|X - 18| < 10\} \\
 &= P\{|X - \mu| < \varepsilon\} \\
 &\geq 1 - \frac{\sigma^2}{\varepsilon^2} \\
 &= 1 - \frac{2.5^2}{10^2} = 0.9375.
 \end{aligned}$$

◦ 马尔可夫不等式

**Example.**  $X_1, X_2, \dots, X_n : i.i.d, X \sim N(\mu, \sigma^2)$ , 证明:

1.  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
2. 设  $Y_i = \frac{X_i - \mu}{\sigma}, i = 1, 2, \dots, n$  则  $E\left(\sum_{i=1}^n Y_i^2\right) = n$

证明. 1. 由线性性:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(E\bar{X}, D\bar{X}).$$

由于  $X$  之间相互独立, 有  $D(X_1 + X_2) = D(X_1) + D(X_2)$

$$E\bar{X} = \frac{1}{n} \sum_{i=1}^n EX_i = \mu, \quad D\bar{X} = \frac{1}{n^2} \sum_{i=1}^n DX_i = \frac{\sigma^2}{n}.$$

2. 由题:  $EY_i = 0, DY_i = 1$

$$E\left(\sum_{i=1}^n Y_i^2\right) = \sum_{i=1}^n EY_i^2.$$

**Notation.**  $Y_i^2$  符合自由度为 1 的卡方分布:  $Y_i^2 \sim \chi^2(1)$

即:  $\sum_{i=1}^n E(Y_i^2) = nE(Y_i^2)$

由方差的定义:  $D(Y_i) = E(Y_i^2) - E(Y_i)^2$  :

$$EY_i^2 = D(Y_i) + E(Y_i)^2 = 1 + 0^2 = 1$$

$$\sum_{i=1}^n E(Y_i^2) = nE(Y_i^2) = n.$$

□

#### 4.4 协方差的性质

- $\text{cov}(X, Y) = \text{cov}(Y, X)$  (对称性)
- $\text{cov}(aX, bY) = ab\text{cov}(X, Y)$

证明. 已知:  $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

$$\begin{aligned} \text{cov}(aX, bY) &= E(aXbY) - E(aX)E(bY) \\ &= abE(XY) - abE(X)E(Y) \\ &= ab\text{cov}(X, Y). \end{aligned}$$

□

- $\text{cov}(c, X) = 0$

**Notation.** 协方差用于衡量随机变量之间的线性关系, 常数和其他随机变量不存在线性关系

证明.

$$\begin{aligned}\operatorname{cov}(cX) &= E(cX) - E(c)E(X) \\ &= cE(X) - cE(X) \\ &= 0.\end{aligned}$$

□

**Notation.**  $\operatorname{cov}(c, c) = D(c) = 0$

$$\circ \operatorname{cov}(aX + bY, cZ) = a\operatorname{ccov}(X + Y) + b\operatorname{ccov}(Y + Z) \quad (\text{分配律})$$

证明.

$$\begin{aligned}\operatorname{cov}(aX + bY, cZ) &= E((aX + bY)cZ) - E(aX + bY)E(cZ) \\ &= E(acXZ + bcYZ) - cEZ(aEX + bEY) \\ &= acE(XZ) + bcE(YZ) - acEXEZ - bcEYEZ \\ &= a\operatorname{ccov}(X, Z) + b\operatorname{ccov}(Y, Z).\end{aligned}$$

□

**Notation.**  $\operatorname{cov}(\sum_{i=1}^n a_i X_i, b_i Z) = \sum_{i=1}^n a_i b_i \operatorname{cov}(X_i, Z)$

**Notation.**  $D(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 DX_i + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j \operatorname{cov}(X_i, X_j)$

## 4.5 相关系数

### 4.5.1 标准化

$$X^* = \frac{X - EX}{\sqrt{DX}}.$$

标准化后的变量  $EX^* = 0, DX^* = 1$

**Definition.**  $X^*, Y^*$  的协方差  $\operatorname{cov}(X^*, Y^*)$  为  $X, Y$  的相关系数  $\rho(X, Y)$

$$\begin{aligned}\operatorname{cov}(X^*, Y^*) &= \operatorname{cov}\left(\frac{X - EX}{\sqrt{DX}}, \frac{Y - EY}{\sqrt{DY}}\right) \\ &= \frac{1}{\sqrt{DX}\sqrt{DY}} \operatorname{cov}(X - EX, Y - EY).\end{aligned}$$

易得  $\operatorname{cov}(X - EX, Y - EY) = \operatorname{cov}(X, Y)$

$$\begin{aligned}\operatorname{cov}(X^*, Y^*) &= \frac{\operatorname{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} \\ &= \rho(X, Y).\end{aligned}$$



## 4.5.2 性质

- $|\rho(X, Y)| \leq 1$
- $P\{X^* = \pm Y^*\} = 1$  是  $\rho(X, Y) = \pm 1$  的充要条件

## Lecture 12

10.29

**Notation.** 相关系数/Pearson 相关系数: 描述两个随机变量之间的线性相关性  
只能描述数值性的变量

$|\rho(X, Y)| = 1$  时: 正相关

$|\rho(X, Y)| > 0.8$ : 强相关

$|\rho(X, Y)| \in (0, 0.5)$ : 弱相关

$\rho = 0$ : 不相关/非线性关系

**Notation.** 相关系数本质上描述:

$$P\{Y = aX + b\}.$$

**Example.**  $f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & x^2 + y^2 > 1 \end{cases}$ , 求:

1.  $X, Y$  的相关性; 2.  $X, Y$  的独立性

解: 1.

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y) dx dy \\ &= \int_{-1}^1 dx \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{x}{\pi} dy \\ &= 0. \end{aligned}$$

同理  $EY = 0$ , 即不相关

2.

$$\begin{aligned} f_X(x) &= \int_D f(x, y) dy \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy \\ &= \frac{2}{\pi} \sqrt{1-x^2}. \end{aligned}$$

同理  $f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2}$ , 易得  $f(x, y) \neq f_X(x) f_Y(y)$ , 即不独立

## 数理统计部分

### 5 大数定律和中心极限定理

**Definition.** 大数定律:

$$\bar{X} \xrightarrow[P]{n \rightarrow +\infty} EX.$$

即: 以某事件发生的频率估计该事件的概率

**Definition.** 中心极限定理:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

其中  $X_1, X_2, \dots, X_i$  独立同分布

该随机变量序列存在分布, 中心极限定理提出不论  $\bar{X}$  的分布是什么, 该序列的分布为正态分布

$$\bar{X} \xrightarrow[n \rightarrow \infty]{L} N(E\bar{X}, D\bar{X}).$$

如何判断随机变量的敛散性:

**Corollary.** 依概率收敛:

对  $\forall \varepsilon$  有:

$$\lim_{n \rightarrow \infty} P\{|X_n - X| < \varepsilon\} = 1.$$

代表序列  $\{X_n\}$  收敛于随机变量  $X$ , 记为  $X_n \xrightarrow[n \rightarrow \infty]{P} X$

**Corollary.** 依分布收敛:

序列的分布函数为  $F_n(x)$ , 随机变量的分布函数  $F(x)$ , 对  $\forall x$ , 有:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

则  $\{X_n\}$  依分布收敛于  $X$ , 记为  $X_n \xrightarrow[n \rightarrow \infty]{L} X$

**Notation.** 测度变换: 通过将问题映射到另一个空间简化计算

依分布收敛要求更弱, 即: 依概率收敛  $\Rightarrow$  依分布收敛

当收敛对象为常数时二者可互推

**Notation.** 撞骗: 只要发出的短信足够多, 成功率符合大数定律

三大大数定律:

$$\left\{ \begin{array}{l} \text{切比雪夫大数定律: 最根本} \\ \text{伯努利大数定律: 例子} \\ \text{辛钦大数定律} \end{array} \right. .$$

## 5.1 大数定律

**Notation.** 切比雪夫大数定律:

**Definition.**  $\{X_i\}$  *i.i.d*,  $\exists EX_i, DX_i$ , 且  $\exists C$ , 使得  $DX_i \leq C$  (方差有界), 则对  $\forall \varepsilon > 0$  当:

$$\lim_{n \rightarrow \infty} P\{|\overline{X_n} - E\overline{X_n}| < \varepsilon\} = 1.$$

时:

$$\overline{X_n} \xrightarrow[n \rightarrow +\infty]{P} E\overline{X_n}.$$

证明.  $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ , 有:

$$\begin{aligned} E\overline{X_n} &= \frac{1}{n} \sum_{i=1}^n EX_i \\ D\overline{X_n} &= \frac{1}{n^2} \sum_{i=1}^n DX_i \\ &\leq \frac{C}{n}. \end{aligned}$$

由切比雪夫不等式:

$$\begin{aligned} P\{|\overline{X_n} - E\overline{X_n}| < \varepsilon\} &\geq 1 - \frac{D\overline{X_n}}{\varepsilon^2} \\ &\geq 1 - \frac{C}{n\varepsilon^2}. \end{aligned}$$

当  $n \rightarrow \infty$  时原式收敛于 1

□

**Notation.** 辛钦大数定律: 序列中的随机变量独立同分布

**Notation.** 伯努利大数定律: 序列中  $X_i \sim B(1, p)$  (已知分布), 记  $\mu_s$  为随机变量序列之和, 有:

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{\mu_s}{n} - p\right| < \varepsilon\right\} = 1.$$

即:  $\frac{\mu_s}{n}$  依概率收敛于  $p$

## 5.2 中心极限定理

**Example.** 高尔顿钉板

**Corollary.** *i.i.d* 的中心极限定理:

$$\lim_{n \rightarrow \infty} P\left\{\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \leq x\right\} = \Phi(x).$$

**Corollary.** 棣莫弗-拉普拉斯定理:  $X_i$  独立同分布,  $X_i \sim B(1, p)$ , 令  $Y = \sum_{i=1}^n X_i$ , 对  $\forall x$  有:

$$\lim_{n \rightarrow \infty} P\left\{\frac{Y - np}{\sqrt{np(1-p)}} \leq x\right\} = \Phi(x).$$

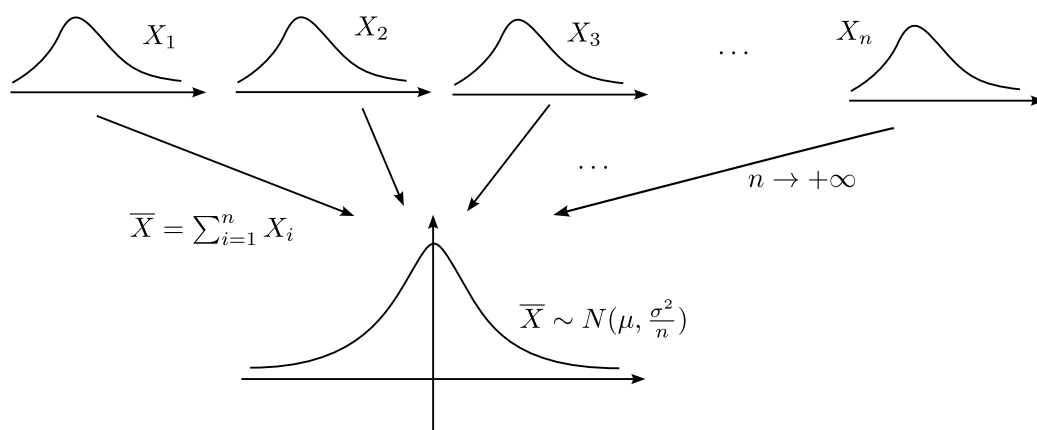


图 6: 中心极限定理

## Lecture 13

10.31

**Notation.** 偏度  $r_1$ : 三阶标准化随机变量的矩, 用于描述对称性

峰度  $r_2$ : 四阶标准化随机变量的矩, 一般使正态分布的峰度  $r_2 = 0$ , 描述分布的陡峭程度

表 3: 常见分布的数字特征

分布	$EX$	$DX$	$r_1$	$r_2$
$B(1, p)$	$p$	$p(1-p)$	$\frac{1-2p}{\sqrt{p(1-p)}}$	$\frac{1}{p(1-p)-6}$
$B(n, p)$	$np$	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$	$\frac{1-6p(1-p)}{np(1-p)}$
$P(\lambda)$	$\lambda$	$\lambda$	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$
$G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{2-p}{\sqrt{1-p}}$	$6 + \frac{p^2}{1-p}$
$U[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	0	$\frac{9}{5} - 3$
$\Gamma(1, \lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2	6
$N(\mu, \sigma^2)$	$\mu$	$\sigma^2$	0	0

## 6 数理统计基本概念

随机变量引入: 使样本空间映射到实数轴上

分布函数: 任意随机变量的概率

大数定律和中心极限定理：由概率论过渡到数理统计

$$\left\{ \begin{array}{l} \text{描述统计学：过去的实验数据/相关分析图} \\ \text{推断统计学：根据现有的实验数据决策} \end{array} \right\} \left\{ \begin{array}{l} \text{参数估计：第七章} \\ \text{假设检验：第八章} \\ \text{回归分析：第九章} \end{array} \right.$$

**Definition.** 总体：全部研究对象，可以用分布描述（随机变量组）

**Definition.** 个体：组成总体的成员，符合总体分布（每一个个体都是一个随机变量）

**Example.** 从总体中抽取  $n$  个样本

对数据记录： $x_1, x_2, \dots, x_n$  称为  $n$  维随机变量  $X_1, X_2, \dots, X_n$  对应的观测值， $X_1, X_2, \dots, X_n$  为来自总体  $X$  的一个样本

**Notation.** 简单样本： $X_1, X_2, \dots, X_n$  *i.i.d.*，且与总体分布相符

特点：

- 独立性
- 代表性

**Definition.** 样本空间： $\Omega = \{(x_1, x_2, \dots, x_n) | x_i \in \mathbb{R}, i = 1, 2, \dots, n\}$

**Notation.** 样本联合分布和总体分布的关系 (*i.i.d.*):

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\} \\ &= \prod_{i=1}^n P\{X_i \leq x_i\} \\ &= \prod_{i=1}^n F(x_i). \end{aligned}$$

扩展： $X$  为连续型，密度函数的关系：

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n f_{X_i}(x_i) \\ &= \prod_{i=1}^n f(x_i) \quad x_i \in \mathbb{R}, i = 1, 2, \dots, n. \end{aligned}$$

## 6.1 经验分布函数

经验分布函数： $F_n(x)$

将样本观测值  $x_1, x_2, \dots, x_n$  按大小分类为  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$

$$\begin{aligned} F_n(x) &= f_n\{X \leq x\} \\ &= \begin{cases} 0, & x < x_{(1)} \\ \frac{k}{n}, & x \in [x_{(k)}, x_{(k+1)}) \\ 1, & x \geq x_{(n)} \end{cases} \\ &\approx F(X). \end{aligned}$$

**Corollary.** 格利文科定理:

$$P \left\{ \limsup_{n \rightarrow \infty, x \in \mathbb{R}} |F(x) - F_n(x)| = 0 \right\} = 1.$$

根据格利文科定理: 可以使用经验分布函数来估计理论分布函数

## Lecture 14

11.05

### 6.2 密度函数

**Notation.** 密度函数和分布函数的关系:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{+\infty} f_X(x) dx \\ f_X(x) &= \frac{dF_X(x)}{dx}. \end{aligned}$$

对于直方图: 将中点光滑连接 = 密度函数

或: 核密度

#### 直方图

**Notation.** 直方图的面积代表频率:

$$\text{高度} h_i = \frac{\text{面积} f_i}{\text{区间长度} \Delta x_i}.$$

直方图的高度代表密度, 直方图的横坐标的取值范围为观测值的取值范围, 直方图分块的区间来源一般为经验公式:  $m \approx 1.87(n-1)^{0.4}$ , 其中  $m$  为区间分组数量

计算直方图频率:

$$\text{频率} f_i = \frac{\text{落入区间的个数} y_i}{\text{总个数} y}.$$

### 6.3 统计量

统计量 (statistic), 统计学 (statistics)

**Definition.** 统计量: 关于样本的函数, 不含任何未知参数

完整定义:

**Example.**  $X_1, X_2$  来自正态总体  $N(\mu, \sigma^2)$  的样本 (这两个任意抽出一个都属于一个样本), 其中  $\mu, \sigma$  均未知, 以下表达式:

- $\frac{1}{4}(X_1 + X_2) - \mu$
- $\frac{X_1}{\sigma}$

均不是统计量 (使用了未知的数), 以下表达式都是统计量

- $3X_1$
- $X_1 - 8$
- $X_1^2 + X_2^2$

提出统计量的目的: 通过样本估计或检测未知量, 因此统计量不能含未知量

常见统计量:

- 样本均值 (算术平均数):  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- 样本方差:  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$
- 样本标准差:  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$
- 样本阶原点矩
- 样本阶中心矩

**Notation.** 样本均值: 若  $X_1, X_2, \dots, X_n$  i.i.d: 根据辛钦大数定律:  $\bar{X} \xrightarrow[n \rightarrow +\infty]{P} E\bar{X} = EX$

**Notation.** 样本阶中心矩:

$$B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \xrightarrow[n \rightarrow \infty]{P} DX.$$

或:  $S^2 = B_2 \times \frac{n}{n-1} \Rightarrow E(S^2) = DX$

证明.

$$\begin{aligned}
 S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \\
 &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\
 &= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 \right) - \sum_{i=1}^n \bar{X}^2 \\
 ES^2 &= E \left[ \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \right] \\
 &= \frac{1}{n-1} \left( \sum_{i=1}^n EX_i^2 - nE\bar{X}^2 \right) \\
 &= \frac{1}{n-1} \left[ \sum_{i=1}^n (DX_i + (EX_i)^2) - n(D\bar{X} + (E\bar{X})^2) \right] \\
 &= \frac{1}{n} \left[ \sum_{i=1}^n (DX + (EX)^2) - n \left( \frac{DX}{n} + (EX)^2 \right) \right] \\
 &= \frac{1}{n-1} [nDX + n(EX)^2 - DX - n(EX)^2] = \frac{1}{n-1} (n-1)DX = DX.
 \end{aligned}$$

即:  $EB_2 = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}DX$

□

用样本均值估计总体均值:

$$\sum_{i=1}^n (X_i - \bar{X})^2 \leq \sum_{i=1}^n (X_i - x)^2.$$

### 顺序统计量

令  $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$  为最小顺序统计量, 最大同理

要求第几小的顺序统计量:  $R$  成为样本极差,  $\tilde{X}$  称为样本中位数

## 6.4 样本均值的分布

**Theorem.**  $X_1, X_2, \dots, X_n$  来自  $N(\mu, \sigma^2)$ , 则

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

定义  $\bar{X}$  为  $X_1, X_2, \dots, X_n$  的线性函数,  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ , 计算期望和方差, 将  $\bar{X}$  标准化

**Theorem.** 标准化后的线性函数  $\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$ :

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \xrightarrow[n \rightarrow \infty]{L} N(0, 1).$$



**Example.** 总体:  $X \sim N(20, 9)$ , 求样本容量  $n$  多大时使样本均值与总体均值的绝对值之差  $\leq 0.3$  的概率  $> 95\%$

### 6.4.1 三大抽样分布

- 卡方分布:  $\chi^2(n)$

**Notation.** 卡方分布实际上为  $\alpha = \frac{1}{2}, \lambda = \frac{n}{2}$  的 Gamma 分布

当  $n = 2$  时为参数为  $\frac{1}{2}$  的指数分布

一般称  $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$  为伽马分布族

**Definition.** 设  $X_1, X_2, \dots, X_n \sim N(0, 1) i.i.d$ , 令  $\chi^2 = \sum_{i=1}^n X_i^2$ , 称  $\chi^2$  为自由度为  $n$  的卡方分布

**Notation.** 卡方分布具有可加性:

$$Y_1 \sim \chi^2(m), Y_2 \sim \chi^2(n) : Y_1 + Y_2 \sim \chi^2(m+n).$$

从  $n = 3$  开始, 卡方分布出现最大值, 且  $n$  越大卡方分布的方差越大  
卡方分布的性质:

- $E(\chi^2) = n, D(\chi^2) = 2n$
- 可加性
- 分位点:

对性质 1:

证明.

$$\begin{aligned} E\left(\sum_{i=1}^n X_i^2\right) &= \sum_{i=1}^n E(X_i^2) = nEX^2 \\ &= n(DX + (EX)^2) \\ D\left(\sum_{i=1}^n X_i^2\right) &= nDX^2 = n(E(X^2)^2 - (EX^2)^2) \\ &= n(EX^4 - 1). \end{aligned}$$

□

## Lecture 15

11.07

## Lecture 16

11.12

*Review:*

1. 抽样分布定理: 定理 6.4.3
2. 表达式  $\frac{(n-1)S^2}{\sigma^2}$  什么时候是统计量:  $\sigma^2$  已知  
当  $\sigma^2$  未知时: 表达式符合  $\chi^2(n-1)$ , 称为**枢轴量**

**Corollary.**  $X_1, X_2, \dots, X_n$  来自总体  $X \sim N(\mu, \sigma^2)$ , 样本均值和方差记为  $\bar{X}, S^2$ , 则:

- a. 求  $ES^2$ :  $ES^2 = \sigma^2$
- b. 求  $DS^2$ :  $DS^2 = \frac{2\sigma^4}{n-1}$
- c. 构造  $t$  分布:  $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}/\sqrt{\frac{S^2}{\sigma^2}} = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)$

(上述  $\sqrt{(n-1)S^2/\sigma^2} \sim \chi^2(n-1)$  且  $(\bar{X}-\mu)/\frac{\sigma}{\sqrt{n}} \sim \Phi(x)$ )

重点题目: 例 6.4.4

**Corollary.** 两个正态总体的抽样分布定理: 两个总体记为  $X, Y$ , 样本容量分别为  $m, n$ , 样本分布分别符合  $\mu = \mu_1, \sigma^2 = \sigma_1^2$  和  $\mu = \mu_2, \sigma^2 = \sigma_2^2$  的正态分布, ...

## 7 参数估计

参数: *param*

- 点估计
  - 矩估计
  - 极大似然估计
- 区间估计

### 7.1 矩估计

使用样本矩  $\bar{X}$  替代总体矩  $\hat{\mu}$

**Notation.** 总体矩不存在 (无穷) 时不能使用矩估计

**Example.** 总体  $X \sim U[a, b]$ ,  $X_1, X_2, \dots, X_n$  为总体的样本, 求  $a, b$  的矩估计量

解:

$$\begin{cases} EX &= \frac{a+b}{2} \\ DX &= \frac{(b-a)^2}{12} \end{cases}.$$

求解得:  $\begin{cases} a &= EX - \sqrt{3DX} \\ b &= EX + \sqrt{3DX} \end{cases}$ , 替代后为:  $\begin{cases} \hat{a} &= \bar{X} - \sqrt{3M_2^*} \\ \hat{b} &= \bar{X} + \sqrt{3M_2^*} \end{cases}$ ,  $\hat{a}$  代表对  $a$  的估计

## 7.2 极大似然估计

似然函数: 样本的联合概率分布函数 (P167)

**Example.** 同矩估计例题, 求  $a, b$  的极大似然估计量

解: 写出联合密度函数:  $L(\dots) = \prod_{i=1}^n f_{X_i}(x_i)$ , 由于  $f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{Others} \end{cases}$ , 则联合密度函数为:

$$\begin{aligned} L &= \prod_{i=1}^n \frac{1}{b-a} I_{[a,b]}(x_i) \\ &= \frac{1}{(b-a)^n} \prod_{i=1}^n I_{[a,b]}(x_i). \end{aligned}$$

## Lecture 17

11.14

## Lecture 18

11.19

*Review:*

- 对参数 (完全可观测数据) 进行点估计: 极大似然估计、矩估计  
两种方法得到的结果可能一样或不一样
- 点估计的评价标准:

(渐进) 无偏 估计的参数的期望  $E\hat{\theta}$  求极限为  $\theta$

有效性 在无偏的前提下:  $D\hat{\theta}$  越小越有效

相合性 (一致性)  $\text{MSE}(\hat{\theta}, \theta)$ : 均方误差

**Example.**  $\bar{X}$  和  $\hat{W} = \sum_{i=1}^n a_i X_i$  可以证明都是  $\mu$  的无偏估计, 称  $\bar{X}$  为算术均值,  $\hat{W}$  为加权均值, 且算术均值比加权均值更有效 (均值不等式)

**Notation.** 均方误差:  $\text{MSE}(\hat{\theta}, \theta) = E(\hat{\theta} - \theta)^2$

## 7.3 区间估计

**Definition.** 置信区间:  $\alpha$  为给定值, 总体的分布函数为  $F(x, \theta)$ , 有两个从总体中抽取后构造的统计量  $T_1, T_2$ , 当:

$$P\{T_1 < \theta < T_2\} = 1 - \alpha.$$

时: 称  $P$  为置信度, 区间长度  $T_2 - T_1$  的数学期望  $E(T_2 - T_1)$  为精度

纽曼提出的准则: 先确定  $\alpha$  来确定置信度, 再确定置信上下限

**Notation.** 已知标准正态分布  $X \sim N(\mu, \sigma^2)$  中的  $\sigma^2$ , 置信度为  $1 - \alpha$  时参数  $\mu$  的置信区间:

$$(T_1, T_2) = \left( \bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \right).$$

如果  $\sigma^2$  未知: 通过  $t$  分布可得:  $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$ , 置信区间为:

$$(T_1, T_2) = \left( \bar{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) \right).$$

**Notation.** 卡方分布下: 方差为  $\sigma^2$ , 置信度为  $1 - \alpha$  的置信区间:

$$(T_1, T_2) = \left( \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} \right).$$

对应的标准差为  $\sigma$ , 置信度为  $1 - \alpha$  的置信区间:  $T'_1 = \sqrt{T_1}, T'_2 = \sqrt{T_2}$

如果  $\mu$  已知,  $\sigma^2$  未知, 令  $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ , 因为  $\chi^2 = \frac{nS_1^2}{\sigma^2} \sim \chi^2(n)$ , 可得置信度为  $1 - \alpha$  的方差的置信区间为:

$$\left( \frac{nS_1^2}{\chi_{1-\frac{\alpha}{2}}^2(n)}, \frac{nS_1^2}{\chi_{\frac{\alpha}{2}}^2(n)} \right).$$

## 8 假设检验

**Notation.** 在医药研发大量应用

- 参数假设检验: 假设效果
- 非参数假设检验

**Notation.** 下节课之前准备一个本专业的假设检验问题

### Lecture 19

11.21

**Example.** 参数假设检验: 主场优势: NBA 某球队进行 82 场比赛, 41 场为主场, 统计过去 15 年主场胜率  $P_1$  和客场胜率  $P_2$ , 判断该球队是否存在主场优势

通过胜率差  $\Delta P = P_1 - P_2$ , 当  $\Delta P = 0$  时不存在, 当  $\Delta P > 0$  时存在

**Notation.** 非参数假设检验: 住房面积和家庭生活幸福感的关系, 学历程度和年均收入的关系; 非参数假设检验的两个变量独立

**Example.** 规定工业废水中 Cr(VI) 排放浓度不超过 0.5, 即  $c_{\text{Cr(VI)}} \leq 0.5$ , 设  $X$  为工业废水有害物质排放浓度总体, 抽取 16 份废水  $X_1, X_2, \dots, X_{16}$ , 测得物质浓度  $\bar{x} = 0.52, s^2 = 0.09$ , 假设该物质浓度分布为  $X \sim N(\mu, \sigma^2)$ , 求排放浓度是否符合规定 ( $\alpha = 0.5$ )

解:

$$\bar{X} = \frac{1}{16} \sum_{i=1}^{16} X_i \quad S^2 = \frac{1}{15} \sum_{i=1}^{16} (X_i - \bar{X})^2.$$

判断是否超标: 通过检测  $\mu \leq 0.5$  达标,  $\mu > 0.5$  超标: 一般把带等号的假设设为原假设  $H_0: \mu \leq 0.5$ , 被则假设  $H_1: \mu > 0.5$ ;

检验水平  $\alpha = 0.05$ : 通常不超过 0.1, 表示犯第一类错误的概率 ( $\alpha = P(\overline{H_0})$ ); 对比  $\beta$ : 犯第二类错误

选择检验统计量: 将最大似然估计标准化:  $\frac{\bar{X} - \mu}{s/\sqrt{n}} = T$ , 当  $H_0$  成立时:  $T = \frac{\bar{X} - 0.5}{S/\sqrt{n}} \sim t(15)$  (分布为  $t$  分布, 该检验方法为  $t$  检验法)

确定拒绝域  $\mathcal{R}_0$ : 确定拒绝域的形式, 由于被则假设为  $H_1: \mu > 0.5$ , 假设拒绝域为  $\{\bar{X} - 0.5 > c\}$ ,  $c$  为未知量

犯第一类错误的概率  $P\{\bar{X} - 0.5 > c \mid \mu \leq 0.5\} \leq 0.05$ , 原式放缩:

$$\begin{aligned} P\{\bar{X} - 0.5 > c \mid \mu \leq 0.5\} &= P\left\{\frac{\bar{X} - 0.5}{\frac{S}{\sqrt{16}}} > \frac{c}{\frac{S}{\sqrt{16}}} \mid \mu \leq 0.5\right\} \\ &= P\left\{\frac{\bar{X} - \mu}{\frac{S}{4}} > \frac{c + 0.5 - \mu}{\frac{S}{4}} \mid \mu \leq 0.5\right\} \\ &\leq P\left\{\frac{\bar{X} - \mu}{\frac{S}{4}} > \frac{c}{\frac{S}{4}}\right\} \\ &= P\left(T > \frac{c}{S/4}\right) = \alpha \quad (0.5 - \mu \geq 0). \end{aligned}$$

根据  $t$  分布的分位数定义可得:  $\frac{c}{S/4} = t_{0.95}(15)$ , 则检验统计量在拒绝域中可以表示为:

$$\mathcal{R} = \left\{\frac{\bar{X} - 0.5}{S/4} > t_{0.95}(15)\right\}.$$

查表得  $t_{0.95}(15) = 1.753$ , 判断:  $\bar{x} = 0.52, s = 0.3$ , 带入原式得:  $\frac{\bar{x} - 0.5}{s/4} = \frac{0.52 - 0.5}{0.3/4} < 1.753$ , 因此并未落在拒绝域中, 接受原假设  $H_0$

下结论: 认为  $t \notin \mathcal{R}_0$ , 因此不拒绝原假设  $H_0$ , 在显著性水平  $\alpha = 0.05$  下, 可以认为该区域有害物质排放浓度符合规定

### 假设检验的两类错误

假设  $H_0$  正确, 可以认为  $H_0$  正确或错误, 类似二分类问题

表 4: 假设检验的两类错误

真实情况 \ 判断	接受 $H_0$	拒绝 $H_0$
$H_0$ 为真	正确	第一类错误 $\alpha$
$H_0$ 为假	第二类错误 $\beta$	正确

对于上题:  $\beta = \left\{ \frac{X-0.5}{s/4} \leq t_{0.95}(15) \mid \mu > 0.5 \right\}$

$$\begin{aligned} \beta &= P \left\{ \frac{X-0.5}{\frac{s}{4}} < t_{0.95}(15) \mid \mu > 0.5 \right\} \quad (\text{We assume that: } \mu = 0.6) \\ &= P \left\{ \frac{X-0.5}{\frac{s}{4}} < t_{0.95}(15) + \frac{0.6-0.5}{\frac{s}{4}} \right\} \approx 60\%. \end{aligned}$$

使用被则假设检验一般较大的概率犯错误, 因此一般使用原假设检验  
假设检验的内容:

$$\text{理论依据: 小概率原理} \left\{ \begin{array}{l} \text{参数检验} \left\{ \begin{array}{l} \text{总体 } \mu, \Delta\mu \text{ 的检验} \\ \text{总体 } \sigma, \partial\sigma \text{ 的检验} \end{array} \right. \\ \text{非参数检验} \left\{ \begin{array}{l} \text{分布拟合检验} \\ \text{符号检验} \\ \text{秩和检验: 两个检验的分布是否是同一分布} \end{array} \right. \end{array} \right.$$

**Notation.** 小概率原理: 犯第一类错误的概率  $\alpha$  为小概率; 如果认为  $H_0$  正确, 但是数据观测得到的是  $H_1: \bar{x} - 0.5 > c$ , 则应该反过来认为  $H_1$  正确

## Lecture 20

11.26

**Example.** 假设产品出厂时要求次品率  $p \leq 0.04$ , 从 10000 件产品中抽取 12 件产品发现 3 件次品, 问该批产品能否出厂; 若抽出 1 件次品, 能否出厂

解: 不使用假设检验:  $p_0 = \frac{3}{12} = 0.25 > 0.04$ , 明显不能出厂 (不严谨), 即使  $p_0 = \frac{1}{12} \approx 0.08 > 0.04$  也不能出厂

**Notation.** 问题在于: 12 件产品和 10000 件产品中出现次品数量的分布不一样

使用假设检验: 假设  $p \leq 0.04$ , 抽查率  $\frac{12}{10000} \approx 0.001$ , 认为分布符合  $B(12, 0.04)$ , 即 12 重伯努利试验, 计算得发生事件 “在 12 个产品中抽到 3 个次品的概率为”:

$$P_{12}(3) = C_{12}^3 p^3 (1-p)^9 \approx 0.0097.$$

因此 “在 12 个产品中抽到 3 个次品的概率” 事件是小概率事件, 但一次试验就发生, 因此推翻原假设, 即  $p > 0.04$ , 所以不能出厂

同理,

$$P_{12}(1) \approx 0.306.$$

概率并不算小, 因此可以认为原假设  $p \leq 0.04$  成立

**Notation.** 一致最优势假设检验: 使第二类错误尽可能小

将上述问题建模:

1. 提出假设:  $H_0: p \leq 0.04, H_1: p > 0.04$
2. 样本观测值:  $(x_1, x_2, \dots, x_{12})$
3. 设定小概率阈值:  $\alpha = 0.01$

**Question.** 白糖打包机: 包装得到的糖重量为一个服从正态分布的随机变量  $X \sim N(\mu, \sigma^2)$ , 运行正常时  $\mu = 100$  kg,  $\sigma^2 = 0.05$ , 为检验运行是否正常, 随机抽取 9 袋糖测得净重:

$$(99.3, 98.7, 100.5, 101.2, 98.3, 99.7, 99.5, 102.1, 100.5).$$

求机器是否运作正常 ( $H_0: \mu = 100, H_1: \mu \neq 100$ )

**Solve.** 由于  $\sigma^2 = 0.05$ , 且工艺不改变, 因此可以认为方差不变, 即  $X \sim N(\mu, 0.05)$ ;

假设:  $\mu = \mu_0 = 100$ , 首先求出样本均值和方差:  $\bar{X} = 99.978$

由于小概率事件原理以及:  $\bar{X}$  是  $\mu$  的一致最小方差无偏估计量:  $|\bar{X} - \mu|$  应该很小; 易得  $\bar{X}$  是  $\mu$  的点估计 (矩估计、极大似然估计), 则当  $H_0$  成立时,  $|\bar{X} - \mu_0|$  也应该很小

设定拒绝域, 即当

$$|\bar{X} - \mu_0| > c.$$

时就拒绝原假设

由于  $\alpha = P(\text{犯第一类错误}) = P(\text{拒绝 } H_0 \mid H_0 \text{ 成立})$ , 且  $\bar{X} \sim N(\mu_0, \frac{\sigma}{n})$  (上一章总体的结论), 标准化后为  $U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$

根据标准化后的随机变量转换后:

$$\begin{aligned} P(|\bar{X} - \mu_0| > c) &= P\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} > \frac{c}{\sigma/\sqrt{n}}\right) \\ &= 1 - P\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \leq \frac{c}{\sigma/\sqrt{n}}\right) \\ &= 1 - P\left(\frac{-c}{\sigma/\sqrt{n}} \leq U \leq \frac{c}{\sigma/\sqrt{n}}\right) = \alpha. \end{aligned}$$

这里的显著性水平  $\alpha$  可以随意定, 但一般使用 0.01, 0.025, 0.05 这几个值, 假定  $\alpha = 0.05$ , 则对于标准正态分布:  $P\left(U \leq \frac{c}{\sigma/\sqrt{n}}\right) = \frac{1-\alpha}{2}$ , 查表得到  $u_{\frac{1-\alpha}{2}} = u_{0.475} = \frac{c}{\sigma/\sqrt{n}}$ , 即:

$$c = \frac{u_{0.475} \cdot \sigma}{\sqrt{n}}.$$

已知  $\bar{X} = 99.98, \sigma = \sqrt{0.05}, \mu_0 = 100, \alpha = 0.05$ , 带入后计算可得:

$$|\bar{X} - \mu_0| = 0.02.$$

**假设检验的种类**

- a.*  $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$ : 双侧检验 (简单原假设/简单备择假设)
- b.*  $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$ : 右侧检验 (拒绝域在右边)
- c.*  $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$ : 左侧检验 (拒绝域在左边)

**Notation.** 期末考试规范性: 10 分, 每步 2 分

- a.* 提出假设  $H_0, H_1$
- b.* 选择统计检验量  $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$  or  $\frac{\bar{X}-\mu}{S/\sqrt{n}}$
- c.* 确定拒绝域  $\mathcal{R}$
- d.* 计算统计检验量的样本值, 观察是否在拒绝域内
- e.* 下结论

**Example.** 书上例题 8.2.1: 慢性铅中毒