

概率论与数理统计讲义

零基础学习与期末复习

@Codex and XZXZZX

创建时间: 2026 年 7 月 17 日

更新时间: 2026 年 7 月 18 日

讲义介绍

这是一套面向**零基础首次学习**、同时兼顾**期末复习**的概率论与数理统计个人讲义. 它不默认读者已经熟悉符号、术语或公式来源: 新概念先说明要解决的问题和直观含义, 再给出正式定义、适用条件与计算方法.

全书共九章. 第一至第五章建立概率论基础, 包括随机事件、随机变量、多维随机变量、数字特征、大数定律与中心极限定理; 第六章从总体与样本出发, 介绍统计量和抽样分布; 第七至第九章依次讨论点估计、区间估计与假设检验, 形成数理统计的基本推断框架.

每个主要知识单元尽量包含直观理解、正式结论、典型例题、辨析、易错点和小结. 例题不仅展示计算, 还说明为什么选择该方法、关键转折在哪里, 以及结果如何检查. 每章末尾另设章节总结、公式条件、方法选择、证明掌握层级和自测题, 便于集中复习.

建议按以下方式使用:

1. **首次学习:** 顺序阅读正文, 先理解概念和条件, 再跟随例题完成计算, 不跳过每节小结.
2. **做题训练:** 先判断题型和适用条件, 再选择公式; 做完后用量纲、取值范围、概率范围或极端情形检查结果.
3. **期末复习:** 优先回看章末知识结构、公式适用条件、重点辨析和易错点, 再用自测题检查薄弱环节.
4. **证明学习:** A 级内容要求能够独立复现, B 级内容要求理解关键思路, C 级内容掌握准确条件、结论和用途即可.

正文统一使用上分位点记号. 超出普通课程基础要求、主要服务考研的内容会标为“**考研专项**”; 未标注的基础定义、通用公式和常规方法均属于本讲义主线.

目录

讲义介绍	i
第一章 随机事件和概率	1
1.1 本章导读	1
1.2 随机试验、样本空间与随机事件	1
1.2.1 从不确定现象到可研究的试验	1
1.2.2 事件是样本空间的子集	3
1.3 事件的关系与运算	4
1.3.1 把自然语言翻译成集合语言	4
1.3.2 运算律与德摩根律	5
1.4 古典概型与几何概型	6
1.4.1 古典概型: 有限个等可能结果	6
1.4.2 几何概型: 连续区域中的等可能	8
1.5 概率的公理化定义与基本性质	9
1.5.1 频率提供直觉, 公理给出定义	9
1.6 条件概率与乘法公式	11
1.6.1 条件就是缩小后的样本空间	11
1.7 事件的独立性	13
1.7.1 独立不是互斥	13
1.7.2 多个事件的独立性	14
1.8 全概率公式与贝叶斯公式	15
1.8.1 先把样本空间不重不漏地划分	15
1.9 本章总结	19
1.9.1 知识结构与方法选择	19
1.9.2 核心公式及适用条件	19
1.9.3 重点辨析与易错点	19
1.9.4 证明掌握清单	20

1.9.5	自测题	20
第二章	随机变量及其分布	22
2.1	随机变量与分布函数	22
2.1.1	从文字事件到数值变量	22
2.1.2	分布函数: 把全部概率装进一个函数	23
2.2	离散型随机变量与常见分布	27
2.2.1	分布律及其与分布函数的关系	27
2.2.2	二项分布与 0-1 分布	28
2.2.3	几何分布与首次发生问题	30
2.2.4	泊松分布与泊松近似	31
2.2.5	超几何分布: 无放回抽样	32
2.3	连续型随机变量与概率密度	34
2.3.1	密度的含义与基本条件	34
2.3.2	从密度求参数、概率与分布函数	37
2.4	三种常见连续型分布	40
2.4.1	均匀分布	40
2.4.2	指数分布与无记忆性	41
2.4.3	正态分布与标准化	43
2.5	随机变量函数的分布	47
2.5.1	离散型: 合并产生同一函数值的概率	47
2.5.2	连续型: 分布函数法	48
2.5.3	概率积分变换	51
2.6	本章总结	52
2.6.1	知识结构	52
2.6.2	核心公式与适用条件	52
2.6.3	常见题型与方法选择	52
2.6.4	重点辨析	53
2.6.5	易错点	53
2.6.6	证明掌握清单	54
2.6.7	自测题	54
第三章	多维随机变量及其分布	56
3.1	二维随机变量与联合分布函数	56
3.1.1	为什么需要二维随机变量	56

3.1.2	联合分布函数与矩形概率	57
3.1.3	边缘分布函数与独立性	58
3.2	二维离散型随机变量	60
3.2.1	联合分布律、边缘与条件分布	60
3.2.2	离散型独立性	61
3.3	二维连续型随机变量	63
3.3.1	联合密度与区域概率	63
3.3.2	边缘概率密度	65
3.3.3	条件概率密度与条件概率	66
3.3.4	连续型独立性	67
3.4	二维均匀分布与二维正态分布	68
3.4.1	二维均匀分布	68
3.4.2	二维正态分布	69
3.5	二维随机变量函数的分布	70
3.5.1	离散型: 合并产生同一函数值的概率	70
3.5.2	连续型通法: 分布函数法	71
3.5.3	和的密度与卷积	72
3.5.4	正态线性组合	73
3.6	最大值、最小值与可加性	74
3.6.1	独立随机变量的最大值与最小值	74
3.6.2	常见可加性结论	75
3.7	本章总结	76
3.7.1	知识结构	76
3.7.2	核心公式与适用条件	76
3.7.3	常见题型与方法选择	76
3.7.4	重点辨析	77
3.7.5	易错点	77
3.7.6	证明掌握清单	77
3.7.7	自测题	78
第四章	随机变量的数字特征	79
4.1	数学期望	79
4.1.1	定义、存在性与直观含义	79
4.1.2	随机变量函数的期望	81
4.1.3	二维随机变量函数的期望	81

4.1.4	期望的运算性质	83
4.2	方差与标准差	83
4.2.1	定义与计算公式	83
4.2.2	方差的运算性质	85
4.3	常见分布的期望与方差	86
4.3.1	常用结论表	86
4.3.2	独立正态的线性组合	87
4.3.3	最大值与最小值的期望技巧	87
4.4	切比雪夫不等式	89
4.5	协方差	90
4.5.1	定义、计算与性质	90
4.6	相关系数	92
4.7	不相关与独立	95
4.8	本章总结	98
第五章	大数定律和中心极限定理	102
5.1	依概率收敛	102
5.1.1	为什么需要一种新的收敛	102
5.1.2	连续函数下的运算性质	104
5.2	大数定律	104
5.2.1	共同主线: 算术平均稳定到期望	104
5.2.2	切比雪夫大数定律	105
5.2.3	辛钦大数定律	106
5.2.4	两种大数定律的辨析	106
5.2.5	伯努利大数定律与频率稳定性	107
5.2.6	用大数定律处理算术平均	108
5.3	独立同分布中心极限定理	109
5.3.1	从“稳定值”走向“近似分布”	109
5.3.2	列维–林德伯格中心极限定理	110
5.3.3	样本均值形式	110
5.3.4	中心极限定理的解题流程	111
5.4	二项分布的正态近似	113
5.4.1	棣莫弗–拉普拉斯定理	113
5.4.2	什么时候使用正态近似	113
5.5	本章总结	115

第六章 数理统计的基本概念	119
6.1 总体与样本	119
6.1.1 总体不是一堆编号, 而是所研究指标的分布	119
6.1.2 简单随机样本	120
6.1.3 样本的二重性	120
6.2 经验分布函数	121
6.2.1 用样本中“不超过给定值的比例”模仿总体概率	121
6.3 统计量与常用样本统计量	123
6.3.1 统计量: 只加工样本, 不偷用未知信息	123
6.3.2 五类常用统计量	123
6.4 统计量的数字特征	125
6.5 三大抽样分布与上分位点	127
6.5.1 抽样分布与上分位点	127
6.5.2 卡方分布: 独立标准正态的平方和	127
6.5.3 学生分布: 标准正态除以独立卡方均方的平方根	128
6.5.4 F 分布: 两个独立卡方均方之比	129
6.6 一个正态总体的抽样分布	130
6.7 两个正态总体的抽样分布	133
第七章 点估计与估计量的评价	139
7.1 点估计的基本概念	139
7.1.1 从未知参数到样本函数	139
7.2 矩估计法	140
7.2.1 思想来源: 样本矩逼近总体矩	140
7.2.2 一个参数的矩估计	141
7.2.3 两个参数的矩估计	143
7.3 最大似然估计法	143
7.3.1 核心思想: 让已观测数据尽可能“像是会出现的”	143
7.3.2 连续型总体: 对数求导	145
7.3.3 离散型总体: 按取值频数归类	145
7.3.4 分段密度与隐藏约束	146
7.3.5 支持集依赖参数: 最大值在边界	147
7.3.6 两个未知参数与不变性	147
7.4 估计量的评价标准	148
7.4.1 无偏性: 长期平均不系统偏高或偏低	149

7.4.2	有效性: 同为无偏时比较波动	150
7.4.3	一致性: 样本增多时逼近真值	151
第八章	区间估计	155
8.1	置信区间的概念	155
8.1.1	随机的是区间, 不是参数	155
8.1.2	单侧置信限	156
8.2	区间估计的统一构造原理	156
8.2.1	枢轴量	156
8.2.2	母型: 正态总体方差已知时估计均值	157
8.2.3	构造置信区间的五步模板	158
8.3	一个正态总体的置信区间	158
8.3.1	均值的置信区间	158
8.3.2	方差的置信区间	159
8.3.3	一个总体的公式选择表	160
8.4	两个正态总体的置信区间	161
8.4.1	均值差: 两个方差均已知	161
8.4.2	均值差: 两个方差未知但相等	161
8.4.3	方差比	162
8.5	置信区间的精度与样本量	163
8.5.1	置信水平、区间长度和样本量的关系	163
8.5.2	由允许误差反推样本量	163
第九章	假设检验	167
9.1	假设检验的基本思想	167
9.1.1	原假设与备择假设	167
9.1.2	小概率事件原理	168
9.1.3	两类错误与显著性水平	168
9.1.4	标准检验步骤	169
9.2	双侧、单侧检验与 p 值	169
9.2.1	三种拒绝域的统一形式	169
9.2.2	p 值	170
9.2.3	与置信区间的联系	170
9.3	一个正态总体均值的检验	171
9.3.1	方差已知: Z 检验	171

9.3.2	方差未知: t 检验	172
9.4	一个正态总体方差的检验	172
9.5	两个正态总体均值差的检验	174
9.5.1	两个方差已知	174
9.5.2	两个方差未知但相等	174
9.6	两个正态总体方差比的检验	175
9.7	统计结论与数学证明的边界	176
9.7.1	未拒绝不等于证明为真	176
9.7.2	反例与普遍命题	176

第一章 随机事件和概率

1.1 本章导读

概率论研究的不是“完全不知道会发生什么”，而是研究这样一类现象：所有可能结果可以预先说清楚，但在试验发生前，无法确定这一次究竟出现哪个结果。本章先把日常语言中的“可能、同时、至少、条件、互不影响”等说法翻译成事件和概率，再建立后续所有章节都会调用的计算工具。

本章的知识链如下：

随机试验 $\longrightarrow$ 样本空间与事件 $\longrightarrow$ 事件运算 $\longrightarrow$ 概率模型与公理
 $\longrightarrow$ 条件概率 $\longrightarrow$ 独立性 $\longrightarrow$ 全概率与贝叶斯公式。

学习本章时要反复检查三个问题：

1. 题目中的字母事件分别表示什么，事件之间是“或”“且”还是“非”？
2. 使用某个公式前，它的条件是否满足，例如等可能、条件事件概率为正、事件独立或构成完备事件组？
3. 算出的数是否位于 $[0, 1]$ ，是否符合“条件增加后样本空间缩小”等直观判断？

1.2 随机试验、样本空间与随机事件

1.2.1 从不确定现象到可研究的试验

“明天下雨”与“掷骰子得到几点”都不能在事前确定结果，但二者并非毫无规律。概率论首先要把研究对象的范围固定下来。

定义 1.1 (随机试验)

若一个试验满足:

1. 在相同条件下可以重复进行;
2. 每次试验的可能结果不止一个, 而且全部可能结果事先可知;
3. 每次试验前不能确定哪一个结果会出现,

则称它为随机试验, 常记为 E .

第二、三条合起来描述“随机现象”; 再加上第一条, 才能通过大量重复试验研究其稳定规律. 这里的“随机”不等于“结果范围未知”: 骰子的结果不确定, 但所有可能点数仍然是已知的 $1, 2, \dots, 6$.

定义 1.2 (样本点与样本空间)

随机试验 E 的每一个不可再分的可能结果称为**样本点**, 记为 ω ; 所有样本点组成的集合称为**样本空间**, 记为 Ω .

例题 1.3 (写出样本空间)

分别写出下列试验的样本空间:

1. 掷一枚骰子, 记录点数;
2. 观察某视频一小时内的播放量;
3. 记录某种机器的寿命;
4. 同时记录一名学生的身高与体重.

选择方法的原因: 样本空间只回答“所有不可再分的结果有哪些”, 因此先确定一次试验最终记录的是一个数还是一对数.

$$\Omega_1 = \{1, 2, 3, 4, 5, 6\},$$

$$\Omega_2 = \{0, 1, 2, \dots\},$$

$$\Omega_3 = (0, +\infty) \quad \text{或依据实际范围写成} [a, b],$$

$$\Omega_4 = \{(x, y) : x > 0, y > 0\}.$$

结果检查: 一次试验只产生其中一个样本点; 样本空间本身则收集了全部可能结果.

若样本点有限或可列, 称为离散样本空间; 若样本点充满一个区间或区域, 称为连续样本空间. 这个分类会直接发展成第二章的离散型与连续型随机变量.

1.2.2 事件是样本空间的子集

定义 1.4 (随机事件)

样本空间 Ω 的一个子集称为随机事件, 常记为 A, B, C . 一次试验产生的样本点属于 A 时, 称事件 A 发生.

由一个样本点组成的事件称为**基本事件**; 空集 $\emptyset$ 称为**不可能事件**; 整个样本空间 Ω 称为**必然事件**. 不可能事件与必然事件在客观上具有确定性, 但把它们作为随机事件的两个极端情况, 可以使运算规则保持统一.

例题 1.5 (样本点与事件不要混淆)

掷一枚骰子. 令 A 表示“点数为偶数”, 则

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad A = \{2, 4, 6\}.$$

若本次掷出 4, 出现的样本点是 $\omega = 4$, 事件 A 发生. A 不是“这次掷出的结果”, 而是所有能使“偶数”成立的结果集合.

注: “单点概率为零”不等于“该点不可能出现”. 在连续样本空间中, 单个点可能是非空事件, 但其概率仍可为零. 事件是否为空是集合问题, 概率是否为零是度量问题, 二者不能随意反推.

小结

本节把随机现象表示为“样本空间中的一个样本点出现”. 使用时先明确试验记录什么, 再写出 Ω , 最后把题目关心的结果写成 Ω 的子集. 后面所有概率公式都作用在这些事件上.

1.3 事件的关系与运算

1.3.1 把自然语言翻译成集合语言

事件本质上是集合, 所以事件关系就是集合关系; 但做题时还要能把集合式翻译回“发生”的语言.

关系或运算	集合表示	事件语言
包含	$A \subseteq B$	A 发生必然导致 B 发生
相等	$A = B$	两事件含有相同样本点
并	$A \cup B$	A, B 至少一个发生
交	$A \cap B$ 或 AB	A, B 同时发生
差	$A - B = A \cap \bar{B}$	A 发生而 B 不发生
互斥	$AB = \emptyset$	A, B 不能同时发生
对立	$AB = \emptyset, A \cup B = \Omega$	A, B 必有且仅有一个发生

注: 互斥不等于对立. 互斥只说明“两者不能同时发生”, 两者还可能同时不发生; 对立还要求二者并起来覆盖整个样本空间. 对立事件必互斥, 互斥事件未必对立.

例题 1.6 (文字事件的表示)

从一批产品中依次抽取三件, 令 A, B, C 分别表示第一、二、三件为正品. 表示:

1. 第三件为正品;
2. 只有第三件为次品;
3. 恰有一件次品;
4. 至多一件次品.

先把“且”译成交、“或”译成并、“不是”译成对立:

- (1) C ;
- (2) $A\bar{B}\bar{C}$;
- (3) $\bar{A}BC \cup A\bar{B}C \cup AB\bar{C}$;
- (4) $AB \cup AC \cup BC$.

第 (4) 问选择“至少两件正品”来表示, 比逐项枚举“零件或一件次品”更短. 检查时逐

个代入“正正正、次正正”等结果, 确认没有遗漏或重复.

1.3.2 运算律与德摩根律

并与交满足交换律、结合律和分配律:

$$\begin{aligned} A \cup B &= B \cup A, & AB &= BA, \\ A \cup (B \cap C) &= (A \cup B) \cap C, & A(BC) &= (AB)C, \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), & A(B \cup C) &= AB \cup AC. \end{aligned}$$

性质 1.7 (德摩根律)

$$\overline{A \cup B} = \bar{A} \bar{B}, \quad \overline{AB} = \bar{A} \cup \bar{B}.$$

更一般地,

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

注: 证明思路 (B 级): 以第一式为例, 任取样本点 ω . 它属于 $\overline{A \cup B}$, 当且仅当它既不属于 A 也不属于 B , 即同时属于 $\bar{A}$ 与 $\bar{B}$. 两个集合包含完全相同的样本点, 因此相等. 该方法是“逐点判断集合相等”, 比死记“长横变短横”更可靠.

例题 1.8 (用德摩根律识别对立事件)

事件 A 表示“甲产品畅销且乙产品滞销”. 令 B 表示“甲产品畅销”, C 表示“乙产品畅销”, 则

$$A = B\bar{C}, \quad \bar{A} = \overline{B\bar{C}} = \bar{B} \cup C.$$

所以 $\bar{A}$ 表示“甲产品滞销或乙产品畅销”. 关键转折是先把文字写成事件式, 再求整体的对立, 不能只把句中某一个词反过来.

注: 事件的差不是普通代数减法. 优先使用 $A - B = A\bar{B}$, 再用交、并、对立的运算律化简; 普通代数中的“去括号”规则一般不适用于事件差.

小结

本节解决“怎样把复杂文字条件写成事件式”. 先找“且、或、非”, 再检查互斥与对立是否被混淆. 德摩根律负责把“至少一个”“全部”等互为反面的说法连接起来, 后面计算复杂事件概率时会反复使用.

1.4 古典概型与几何概型

1.4.1 古典概型: 有限个等可能结果

定义 1.9 (古典概型)

若样本空间只含有限个样本点, 且每个样本点出现的可能性相同, 则称该试验为古典概型 (等可能概型). 若 $|\Omega| = n$, 事件 A 含 k 个样本点, 则

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

使用古典概型前必须检查“有限”和“等可能”. 若给样本点编号的方式改变, 分子与分母必须始终采用同一种计数口径.

排列与组合是古典概型的计数工具:

$$A_n^m = \frac{n!}{(n-m)!}, \quad \binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

顺序重要时用排列, 顺序不重要时用组合. 组合数恒等式

$$\binom{n}{m} = \binom{n}{n-m}, \quad \binom{n}{m} + \binom{n}{m-1} = \binom{n+1}{m}$$

用于简化计数; 其完整代数推导属于 C 级, 本阶段会解释含义并正确使用即可.

例题 1.10 (生日问题: 正难则反)

假设一年有 365 天且生日等可能, 求 60 人中至少两人生日相同的概率.

选择方法的原因: 正面需要分类“两人相同、三人相同、多个小组分别相同”, 容易重复; 反面“所有人生日互不相同”只有一种结构.

把 60 个人看成不同的球, 把 365 天看成不同的盒子. 全部生日安排有 365^{60} 种; 生日互

不相同有

$$A_{365}^{60} = 365 \cdot 364 \cdots 306$$

种. 因此

$$P(\text{至少两人生日相同}) = 1 - \frac{A_{365}^{60}}{365^{60}} \approx 0.994.$$

结果检查: 人数增加时概率应上升; 答案接近 1 虽然反直觉, 但没有超出概率范围.

例题 1.11 (平均分班中的特殊元素)

15 名新生中有 3 名竞赛生, 随机平均分到 3 个有编号的班级, 每班 5 人. 求:

1. 每班恰有 1 名竞赛生的概率;
2. 3 名竞赛生在同一班的概率.

选择方法的原因: 限制集中在竞赛生身上, 所以先安排特殊元素, 再补普通学生, 可以避免重复和遗漏. 全部分班方法数为

$$\binom{15}{5} \binom{10}{5}.$$

第 (1) 问, 先将 3 名竞赛生全排列到 3 个班, 再分别补入普通学生:

$$P_1 = \frac{3! \binom{12}{4} \binom{8}{4}}{\binom{15}{5} \binom{10}{5}} = \frac{25}{91}.$$

第 (2) 问, 先选择竞赛生所在班, 再补足该班和另一个班:

$$P_2 = \frac{\binom{3}{1} \binom{12}{2} \binom{10}{5}}{\binom{15}{5} \binom{10}{5}} = \frac{6}{91}.$$

检查: 两个事件互斥, 且 $P_1 + P_2 < 1$; 剩余情形是 3 名竞赛生按 2 + 1 + 0 分布.

1.4.2 几何概型: 连续区域中的等可能

定义 1.12 (几何概型)

若样本点充满一个有界区域 Ω , 随机点落入等度量子区域的可能性相同, 则称为几何概型. 事件 A 对应区域 $G \subseteq \Omega$ 时,

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

其中 m 可表示长度、面积或体积.

几何概型把古典概型的“等可能”从有限样本点推广到连续区域. 概率只与区域的度量有关, 与位置和形状无关; 它也是第二章均匀分布的直观背景.

例题 1.13 (二维几何概型)

在区间 $(0, 1)$ 内独立任取两个数 X, Y , 求 $X + Y < \frac{6}{5}$ 的概率.

因为一次试验产生一对数 (X, Y) , 样本空间是单位正方形

$$\Omega = \{(x, y) : 0 < x < 1, 0 < y < 1\}.$$

事件区域位于直线 $x + y = \frac{6}{5}$ 下方. 直接求五边形面积较繁, 改求右上角补三角形: 两条直角边均为 $\frac{4}{5}$. 因此

$$P\left(X + Y < \frac{6}{5}\right) = 1 - \frac{1}{2} \left(\frac{4}{5}\right)^2 = \frac{17}{25}.$$

关键转折是把两个数视为平面坐标; 结果检查是最后必须除以总面积, 本题总面积恰为 1.

注: 经典辨析: 古典概型要求“有限且等可能”, 几何概型要求“连续区域内按度量均匀”. 若只有“随机”二字而没有等可能依据, 不能擅自使用“有利情况数除以总情况数”.

小结

本节把概率计算转化为“有利部分占整体的比例”. 古典概型先判断顺序是否重要, 再计数; 几何概型先建立坐标区域, 再求度量之比. 两者都必须先检查等可能条件.

1.5 概率的公理化定义与基本性质

1.5.1 频率提供直觉, 公理给出定义

进行 n 次相同条件下的试验, 事件 A 发生 n_A 次, 称 n_A 为频数,

$$f_n(A) = \frac{n_A}{n}$$

称为频率. 频率是一次实际试验得到的数; 概率 $P(A)$ 是描述事件发生可能性大小的数学量. 大量重复试验中频率的稳定性将在第五章由大数定律精确说明.

公理 1.14 (概率公理)

设 Ω 为样本空间. 若集合函数 P 满足:

1. 非负性: 对任意事件 A , $P(A) \geq 0$;
2. 规范性: $P(\Omega) = 1$;
3. 可列可加性: 若 $A_1, A_2, \dots$ 两两互斥, 则

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

则称 $P(A)$ 为事件 A 的概率.

概率公理是理论起点, 不再证明 (C 级). 古典概型与几何概型都是满足这三条公理的具体模型.

性质 1.15 (由公理推出的常用性质)

对任意事件 A, B :

$$P(\emptyset) = 0, \quad 0 \leq P(A) \leq 1,$$

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

$$P(A - B) = P(A) - P(AB),$$

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

若 $B \subseteq A$, 则

$$P(A - B) = P(A) - P(B), \quad P(B) \leq P(A).$$

注: 证明思路 (B 级): 所有性质都围绕“把事件拆成两两互斥的部分”. 例如

$$A = (A - B) \cup AB$$

且右侧两部分互斥, 所以由有限可加性得到减法公式; 再将

$$A \cup B = A \cup (B - A)$$

拆成互斥部分, 就得到加法公式. 真正起作用的条件是互斥, 而不是图形长得像面积.

性质 1.16 (有限可加性与容斥公式)

若 $A_1, \dots, A_n$ 两两互斥, 则

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

对任意三个事件,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(AB) - P(AC) - P(BC) + P(ABC). \end{aligned}$$

例题 1.17 (概率为零能推出什么)

已知 $P(AB) = 0$. 能否断定 A, B 互斥?

不能. $P(AB) = 0$ 只说明交事件的概率为零, 不能推出 $AB = \emptyset$. 在连续模型中, AB 可以是一个非空单点事件而概率为零.

但若 $P(AB) > 0$, 则一定有 $AB \neq \emptyset$, 所以 A, B 能同时发生. 这里使用的是“空事件概率必为零”的逆否命题.

例题 1.18 (概率大小比较)

比较 $P(AB), P(A), P(A \cup B), P(A) + P(B)$ 的大小.

利用包含关系

$$AB \subseteq A \subseteq A \cup B$$

得到

$$P(AB) \leq P(A) \leq P(A \cup B).$$

再由加法公式及 $P(AB) \geq 0$,

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

所以

$$P(AB) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B).$$

方法选择: 概率不等式优先找事件包含, 再使用 $0 \leq P(\cdot) \leq 1$.

小结

本节从三条概率公理得到后续计算的基础工具. 使用加法、减法或拆分公式时, 先寻找两两互斥的分块; 遇到概率不等式时, 先找事件包含. 特别注意: 事件关系可以推出相应概率关系, 反向一般不成立.

1.6 条件概率与乘法公式

1.6.1 条件就是缩小后的样本空间

定义 1.19 (条件概率)

设 A, B 为事件且 $P(B) > 0$, 称

$$P(A | B) = \frac{P(AB)}{P(B)}$$

为在事件 B 发生的条件下事件 A 发生的条件概率.

已知 B 发生后, 候选样本点从 Ω 缩小到 B ; 在这个新范围中, 真正使 A 发生的是 AB . 因此条件概率本身仍是概率, 对竖线左侧事件满足非负性、规范性、可加性、对立事件公式等性质.

注: 只能直接化简竖线左侧:

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

而 $P(A | \bar{C})$ 不能写成 $1 - P(A | C)$, 因为右侧是条件, 必须回到定义计算.

定理 1.20 (乘法公式)

若 $P(B) > 0$, 则

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

若 $P(A_1 A_2 \cdots A_{n-1}) > 0$, 则

$$\begin{aligned} P(A_1 A_2 \cdots A_n) &= P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \\ &\quad \cdots P(A_n | A_1 A_2 \cdots A_{n-1}). \end{aligned}$$

证明: 证明目标: 从条件概率定义推出两事件乘法公式.

出发点: 已知 $P(B) > 0$, 所以 $P(A | B)$ 有定义.

关键转换: 把条件概率定义中的分母 $P(B)$ 乘到等式另一边.

正式推导: 由定义

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

因为 $P(B) > 0$, 等式两边同乘 $P(B)$, 得到

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

这就是乘法公式. 多事件情形对 $P(A_1 \cdots A_n)$ 反复应用两事件公式即可.

证明小结: 真正保证变形合法的是 $P(B) > 0$. 这个公式把“一起发生”拆成“先发生条件事件, 再在该条件下发生另一个事件”, 是全概率公式的直接工具. ■

例题 1.21 (不放回拨号)

某人忘记电话号码最后一位, 随机尝试且错误号码不再重复. 求不超过三次拨通的概率.

选择方法的原因: 正面可以分“第 1、2、3 次拨通”, 反面“三次均错误”更容易用乘法公式连乘.

$$P(\text{不超过三次拨通}) = 1 - \frac{9}{10} \cdot \frac{8}{9} \cdot \frac{7}{8} = \frac{3}{10}.$$

每个分数都带有前次失败给出的新信息, 所以不能把三次失败简单当成独立事件. 检查:

十个号码在试验前地位对称, 前三个位置包含正确号码的概率也应为 $3/10$.

小结

本节解决“获得新信息后怎样重算概率”. 先检查条件事件概率是否为正, 再把样本空间缩小到条件事件. 乘法公式适合沿时间或层次逐步展开, 尤其要判断前一步结果是否改变后一步概率.

1.7 事件的独立性

1.7.1 独立不是互斥

定义 1.22 (两个事件相互独立)

若

$$P(AB) = P(A)P(B),$$

则称事件 A 与 B 相互独立. 当 $P(B) > 0$ 时, 这等价于

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

即知道 B 是否发生并不改变对 A 的概率判断.

独立描述“概率互不影响”; 互斥描述“不能同时发生”. 它们讨论的层次不同.

命题 1.23 (正概率互斥事件不独立)

若 A, B 互斥且 $P(A) > 0, P(B) > 0$, 则 A, B 不独立.

证明: 证明目标: 在两个事件概率均为正的前提下, 证明互斥排除独立.

出发点: 互斥给出 $AB = \emptyset$; 独立要求 $P(AB) = P(A)P(B)$.

关键转换: 分别计算交事件概率与两个边缘概率的乘积并比较.

正式推导: 由 $AB = \emptyset$,

$$P(AB) = P(\emptyset) = 0.$$

另一方面, 因为 $P(A) > 0, P(B) > 0$,

$$P(A)P(B) > 0.$$

因此 $P(AB) \neq P(A)P(B)$, 不满足独立定义, 故 A, B 不独立.

证明小结: 正概率条件不可删除. 若 $A = \emptyset, B = \Omega$, 二者既互斥又独立, 这个反例说明没有正概率条件时, 不能声称“互斥必不独立”. ■

性质 1.24 (独立性的常用性质)

若 A, B 独立, 则

$$A \text{ 与 } \bar{B}, \quad \bar{A} \text{ 与 } B, \quad \bar{A} \text{ 与 } \bar{B}$$

也分别独立. 概率为 0 或 1 的事件与任意事件独立.

注: 证明思路 (B 级): 例如

$$P(A\bar{B}) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)P(\bar{B}),$$

故 A 与 $\bar{B}$ 独立. 若 $P(A) = 0$, 由 $AB \subseteq A$ 得

$$0 \leq P(AB) \leq P(A) = 0,$$

于是 $P(AB) = 0 = P(A)P(B)$; 概率为 1 的情形可转化为其对立事件概率为 0.

1.7.2 多个事件的独立性

三个事件 A, B, C 两两独立, 是指

$$P(AB) = P(A)P(B), \quad P(AC) = P(A)P(C), \quad P(BC) = P(B)P(C).$$

它们相互独立还要求

$$P(ABC) = P(A)P(B)P(C).$$

相互独立比两两独立更强; 不能只验证两两乘积就宣称相互独立.

例题 1.25 (至少一人成功)

三人独立猜谜, 猜中的概率分别为 $\frac{1}{5}, \frac{1}{3}, \frac{1}{4}$. 求谜语被猜中的概率.

“至少一人猜中”的反面是“三人都未猜中”. 因为三人的试验独立, 反面事件概率可以相乘:

$$P(\text{至少一人猜中}) = 1 - \frac{4}{5} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{3}{5}.$$

方法检查: 若没有“独立”条件, 最后一步不能直接相乘.

小结

本节解决“一个事件的信息是否会改变另一个事件的概率”. 判断独立的最终标准是乘积等式. 互斥与独立不能混用; 涉及多个事件时, 还必须区分两两独立和相互独立.

1.8 全概率公式与贝叶斯公式

1.8.1 先把样本空间不重不漏地划分

定义 1.26 (完备事件组)

若事件 $B_1, \dots, B_n$ 满足

$$B_i B_j = \emptyset \quad (i \neq j), \quad \bigcup_{i=1}^n B_i = \Omega,$$

则称 $B_1, \dots, B_n$ 构成 Ω 的一个完备事件组, 也称样本空间的一个划分. 每次试验中它们必有且仅有一个发生.

识别这类题的信号是: 题目先给出一组“原因、来源或第一阶段”, 它们不重不漏地覆盖全部情况; 随后研究建立在这些情况之上的结果.

定理 1.27 (全概率公式)

若 $B_1, \dots, B_n$ 为完备事件组, 且 $P(B_i) > 0$, 则对任意事件 A ,

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

证明: 证明目标: 利用样本空间的划分, 把 $P(A)$ 写成各类条件下概率的加权和.

出发点: $\bigcup_{i=1}^n B_i = \Omega$ 且各 B_i 两两互斥.

关键转换: 用 B_i 将 A 分割成互斥事件 AB_i , 再分别使用有限可加性与乘法公式.

正式推导:

$$A = A\Omega = A\left(\bigcup_{i=1}^n B_i\right) = \bigcup_{i=1}^n (AB_i).$$

由于 B_i 两两互斥, AB_i 也两两互斥, 因此

$$P(A) = \sum_{i=1}^n P(AB_i).$$

再由乘法公式 $P(AB_i) = P(B_i)P(A | B_i)$, 得到

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

证明小结: 全概率公式真正使用的是“划分 + 互斥可加 + 乘法公式”. 若分类有重叠或遗漏, 就不能直接套用. ■

定理 1.28 (贝叶斯公式)

在全概率公式的条件下, 若 $P(A) > 0$, 则

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)}.$$

证明: 证明目标: 在已知结果 A 发生后, 反推它来自第 i 类原因 B_i 的概率.

出发点: 条件概率定义、乘法公式和已经证明的全概率公式.

关键转换: 分子 $P(AB_i)$ 用乘法公式展开, 分母 $P(A)$ 用全概率公式展开.

正式推导:

$$P(B_i | A) = \frac{P(AB_i)}{P(A)}$$

$$= \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)}.$$

证明小结: 分子正是全概率展开中的第 i 项, 分母是所有来源产生结果 A 的总概率. 贝叶斯公式不是孤立的新公式, 而是条件概率定义与全概率公式的组合. ■

例题 1.29 (产品来源: 由因到果, 再由果推因)

甲、乙、丙三厂的市场占有率分别为 $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$, 次品率分别为 2%, 1%, 3%. 随机购买一件产品, 求:

1. 买到次品的概率;
2. 已知买到次品, 它来自甲厂的概率.

令 B_1, B_2, B_3 分别表示产品来自甲、乙、丙厂, 令 A 表示产品为次品. 三厂构成完备事件组.

第 (1) 问是“由来源求结果”, 使用全概率公式:

$$\begin{aligned} P(A) &= \frac{1}{4} \times 0.02 + \frac{1}{4} \times 0.01 + \frac{1}{2} \times 0.03 \\ &= 0.0225. \end{aligned}$$

第 (2) 问是“已知结果反推来源”, 使用贝叶斯公式:

$$P(B_1 | A) = \frac{\frac{1}{4} \times 0.02}{0.0225} = \frac{2}{9} \approx 0.2222.$$

检查: 后验概率约 0.2222, 低于甲厂原有份额 0.25, 因为甲厂次品率低于总体平均次品率 2.25%, 方向合理.

例题 1.30 (通信误码)

发送信息 A, B 的频率之比为 3 : 2. A 被误收为 B 的概率为 0.04, B 被误收为 A 的概率为 0.07. 已知接收到 A , 求原发送信息也是 A 的概率.

令 H 表示“原发送 A ”, R 表示“接收到 A ”. 则

$$P(H) = \frac{3}{5}, \quad P(\bar{H}) = \frac{2}{5}, \quad P(R | H) = 0.96, \quad P(R | \bar{H}) = 0.07.$$

先由全概率公式计算收到 A 的概率:

$$P(R) = \frac{3}{5} \times 0.96 + \frac{2}{5} \times 0.07 = 0.604.$$

再由贝叶斯公式反推:

$$P(H | R) = \frac{\frac{3}{5} \times 0.96}{0.604} = \frac{144}{151} \approx 0.9536.$$

关键转折是自己重新设置不混淆的事件字母, 并区分“原发送”和“接收到”两个层次.

性质 1.31 (抽签原理)

在无放回随机抽取中, 试验开始前, 第 k 个位置抽到某一类对象的概率都等于该类对象在总体中所占的比例, 与抽取顺序无关.

注: 抽签原理的完整一般证明属于 C 级, 本阶段不要求复现. 要注意它描述的是试验开始前各位置的边缘概率相同; 一旦前面抽取结果已经公开, 后续位置的条件概率会随信息改变.

小结

本节解决两层随机过程. 全概率公式把“结果”按所有可能原因分解, 贝叶斯公式在观察到结果后更新各原因的概率. 使用时先检查 B_i 是否构成完备事件组, 再写清先验概率 $P(B_i)$ 、条件概率 $P(A | B_i)$ 和所求方向.

1.9 本章总结

1.9.1 知识结构与方法选择

题目特征	优先方法	先检查的条件
有限个等可能结果	古典概型、排列组合	有限、等可能、计数口径一致
连续区域内均匀取点	几何概型	均匀性、区域与度量
至少一个、至多、全都	对立事件、德摩根律	正面还是反面更易算
已知某事件发生	条件概率	条件事件概率大于零
分阶段同时发生	乘法公式	每一步的条件是否改变
概率互不影响	独立性	乘积等式是否成立
原因不重不漏、求结果	全概率公式	原因构成完备事件组
已知结果、反推原因	贝叶斯公式	分母的总概率大于零

1.9.2 核心公式及适用条件

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

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

$$P(AB) = P(B)P(A | B), \quad P(B) > 0,$$

$$A, B \text{ 独立} \iff P(AB) = P(A)P(B),$$

$$P(A) = \sum_i P(B_i)P(A | B_i),$$

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_j P(B_j)P(A | B_j)}.$$

1.9.3 重点辨析与易错点

1. 互斥是交集为空; 独立是概率乘积等式. 正概率互斥事件不独立, 但没有正概率条件时不能直接判断.
2. 互斥未必对立; 对立还要求并集为 Ω .

3. $P(A) = 0$ 未必有 $A = \emptyset$; $P(A) = 1$ 也未必有 $A = \Omega$.
4. 条件概率只对竖线左侧事件使用概率运算律, 右侧条件不能随意取反或拆分.
5. 古典概型的分母不是“看起来所有情况”, 而是与分子采用相同单位划分后的等可能基本结果数.
6. 两两独立不等于相互独立; 多个事件时必须按定义逐级验证.
7. 全概率公式的分类必须不重不漏; 贝叶斯公式的分子是全概率展开中的一项.

1.9.4 证明掌握清单

- A 级: 能从条件概率定义推出乘法公式; 能用事件划分完整推出全概率公式和贝叶斯公式; 能证明正概率互斥事件不独立并给出缺少正概率条件时的反例.
- B 级: 能说明概率基本性质来自“互斥分割”; 能用逐点法解释德摩根律; 能说明独立性对补事件保持以及概率为 0 或 1 的事件为何与任意事件独立.
- C 级: 概率公理作为理论起点不证明; 抽签原理的一般证明和组合恒等式的完整代数推导本阶段不要求复现, 但必须会写条件并正确使用.

1.9.5 自测题

练习 1.32 (概念与事件运算)

设 A, B, C 为事件. 分别表示“恰有两个发生”和“至少两个发生”; 再说明互斥与对立的区别.

练习 1.33 (概率性质)

已知 $P(A) = 0.6, P(B) = 0.5, P(AB) = 0.3$, 求 $P(A \cup B)$ 、 $P(A - B)$ 和 $P(\bar{A}B)$, 并判断 A, B 是否独立.

练习 1.34 (条件概率与独立性)

袋中有 4 个红球、6 个白球, 不放回抽取两次. 求第二次为红球的概率, 并分别用全概率公式和抽签原理解释结果.

练习 1.35 (贝叶斯判断)

某检测对患病者呈阳性的概率为 0.95, 对未患病者误报阳性的概率为 0.02, 人群患病率为 0.01. 求检测阳性者实际患病的概率, 并解释为什么结果不等于 0.95.

本章小结

本章建立了概率论的语言和第一套计算体系: 用样本空间刻画全部可能结果, 用事件运算表达复杂条件, 用概率公理保证计算规则, 再通过条件概率、独立性、全概率公式和贝叶斯公式处理信息、层次与反推问题. 做题时始终先写事件、查条件、再选公式; 这套习惯会直接决定第二章以后能否正确建立随机变量及其分布.

第二章 随机变量及其分布

本章导读

第一章解决了“事件发生的可能性有多大”，但如果每道题都用文字命名事件，知识仍然比较零散。本章引入随机变量，把试验结果数值化，再用分布函数、分布律或概率密度统一记录全部概率信息。后续的数字特征、极限定理和数理统计都建立在这些对象上。

随机试验 $\longrightarrow$ 随机变量 $X \longrightarrow$ 分布函数 F_X ,

$$\longrightarrow \begin{cases} \text{离散型: 分布律 } p_k, \\ \text{连续型: 概率密度 } f_X, \end{cases} \longrightarrow \text{随机变量函数 } Y = g(X).$$

学习本章时始终先问三件事:

1. 研究的随机变量是什么, 它可能取哪些值?
2. 它属于离散型、连续型, 还是两者都不是?
3. 已知的是分布函数、分布律还是密度, 应使用求和、积分还是变量变换?

2.1 随机变量与分布函数

2.1.1 从文字事件到数值变量

定义 2.1 (随机变量)

设随机试验的样本空间为 Ω . 定义在 Ω 上、取实数值的单值函数

$$X = X(\omega), \quad \omega \in \Omega,$$

称为随机变量. 本课程中可直观理解为: 每个试验结果 ω 都被唯一地赋予一个数 $X(\omega)$.

随机变量的“大写”表示取值尚未确定; 小写字母 x 表示一个已经指定的实数. 随机变量不是“会自己变化的普通未知数”, 而是把样本点翻译成数值的函数.

例题 2.2 (三个背景中的随机变量)

1. 从含 5 件废品的 50 件产品中抽取 3 件, 令 X 为废品数, 则 $X \in \{0, 1, 2, 3\}$, “至多两件废品”为 $\{X \leq 2\}$.
2. 购买一台电脑, 令 X 为使用寿命, 则 X 可在某个连续区间内取值, “使用超过五年”为 $\{X > 5\}$.
3. 抛一枚硬币, 规定正面记为 1、反面记为 0, 则 $X \in \{0, 1\}$. 这说明非数值结果也可以通过赋值成为随机变量.

注: 不同的样本点可以对应同一个随机变量值, 但同一个样本点不能对应两个不同值. 对任意 $x_1 \neq x_2$, 事件 $\{X = x_1\}$ 与 $\{X = x_2\}$ 必然互斥.

2.1.2 分布函数: 把全部概率装进一个函数**定义 2.3 (分布函数)**

随机变量 X 的分布函数定义为

$$F_X(x) = P(X \leq x), \quad -\infty < x < +\infty.$$

它表示 X 的取值落在实数 x 左侧 (包含 x) 的概率.

分布函数的自变量必须覆盖整个实数轴. 知道 F_X 后, 任何区间事件的概率都可由端点处的函数值或左极限得到.

定理 2.4 (由分布函数计算区间概率)

记

$$F_X(a-0) = \lim_{x \uparrow a} F_X(x).$$

则

$$\begin{aligned} P(X \leq a) &= F_X(a), & P(X < a) &= F_X(a-0), \\ P(a < X \leq b) &= F_X(b) - F_X(a), & P(a \leq X \leq b) &= F_X(b) - F_X(a-0), \\ P(a < X < b) &= F_X(b-0) - F_X(a), & P(X = a) &= F_X(a) - F_X(a-0). \end{aligned}$$

证明: 证明目标: 从定义推出区间概率与跳跃点概率的公式.

出发点: 分布函数定义、事件包含关系和概率的减法公式.

关键转换: 把目标区间写成两个嵌套事件之差. 例如

$$\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}.$$

正式推导: 因为 $\{X \leq a\} \subseteq \{X \leq b\}$,

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a).$$

又因为

$$\{X = a\} = \{X \leq a\} \setminus \{X < a\},$$

所以

$$P(X = a) = F_X(a) - F_X(a-0).$$

其余开闭区间公式同理, 只需根据端点是否包含选择 $F_X(\cdot)$ 或 $F_X(\cdot-0)$.

证明小结: 真正起作用的是“右端大范围减去左端小范围”. 离散型分布函数可能跳跃, 不能随意把左极限替换为函数值; 连续型情形才可忽略端点开闭. ■

性质 2.5 (分布函数的充要条件)

函数 F 是某个随机变量的分布函数, 当且仅当:

1. F 单调不减;
2. F 右连续;
3. $\lim_{x \rightarrow -\infty} F(x) = 0$;

$$4. \lim_{x \rightarrow +\infty} F(x) = 1.$$

由这些条件还可推出 $0 \leq F(x) \leq 1$.

注: 分段书写分布函数时, 可统一采用“左闭右开”, 即等号跟随 $x \geq a$ 的一段, 以便直接满足右连续. 但这只是排版口诀, 最终仍应以右连续性为判断依据.

左极限公式和充要条件的严格分析证明属于 C 级: 本阶段会准确使用, 不要求用数列极限或海涅定理完整复现.

例题 2.6 (跳跃点与区间概率)

设

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & 0 \leq x < 1, \\ 1 - e^{-x}, & x \geq 1. \end{cases}$$

求 $P(X = 1)$ 与 $P\left(\frac{1}{3} < X < \frac{1}{2}\right)$.

选择方法的原因: 题目直接给出分布函数, 先用端点公式, 不必猜测随机变量类型.

$$P(X = 1) = F(1) - F(1 - 0) = (1 - e^{-1}) - \frac{1}{2}.$$

而 $\frac{1}{3}, \frac{1}{2}$ 同处于分布函数的常数段,

$$P\left(\frac{1}{3} < X < \frac{1}{2}\right) = F\left(\frac{1}{2} - 0\right) - F\left(\frac{1}{3}\right) = \frac{1}{2} - \frac{1}{2} = 0.$$

检查: 第一个结果对应 $x = 1$ 处的向上跳跃量; 常数段内没有概率累积, 所以第二个概率为零.

例题 2.7 (验证一个函数是否为分布函数)

判断

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-x}, & x \geq 0 \end{cases}$$

能否成为分布函数.

验证目标: 逐项检查分布函数的充要条件.

出发点: 单调性、右连续性和两端极限.

关键转换: 分段内部用初等函数性质, 唯一需单独检查的点是 $x = 0$.

逐项验证:

1. 两段均单调不减, 且 $F(0) = 0$, 跨越分段点时没有下降;
2. $\lim_{x \downarrow 0} F(x) = 0 = F(0)$, 故在 0 处右连续; 其余点连续;
3. $\lim_{x \rightarrow -\infty} F(x) = 0$;
4. $\lim_{x \rightarrow +\infty} (1 - e^{-x}) = 1$.

四项全部满足, 因此它是分布函数.

验证小结: 分段函数不能只看取值是否在 $[0, 1]$; 最容易漏掉的是分段点右连续和正负无穷处的极限.

注: [分布函数的简单变换] 若 F, G 都是分布函数, 则:

- $aF + bG$ 在 $a, b \geq 0, a + b = 1$ 时仍是分布函数;
- $F(x)G(x)$ 仍是分布函数;
- $F(ax + b)$ 对 $a > 0$ 一定仍是分布函数; $a < 0$ 会反转单调性, 一般不成立;
- 若 F 连续, 则 $1 - F(-x)$ 是分布函数; 一般情形需用左极限修正为 $1 - F(-x - 0)$.

小结

本节把文字事件统一成随机变量的取值范围, 再用 $F_X(x) = P(X \leq x)$ 收集全部概率信息. 使用时先辨认大写随机变量与小写实数, 再根据端点开闭选择函数值或左极限; 验证分布函数必须检查单调不减、右连续和两端极限.

2.2 离散型随机变量与常见分布

2.2.1 分布律及其与分布函数的关系

定义 2.8 (离散型随机变量与分布律)

若随机变量 X 的全部可能取值为有限个或可列无限多个

$$x_1, x_2, \dots,$$

则称 X 为离散型随机变量. 记

$$P(X = x_k) = p_k,$$

称 $\{p_k\}$ 为 X 的概率分布或分布律.

性质 2.9 (分布律的充要条件)

$$p_k \geq 0, \quad \sum_k p_k = 1.$$

反之, 任何满足这两项的数列都可作为某个离散型随机变量的分布律.

离散型概率计算只有一个核心动作: 找出目标事件包含的所有可能取值, 再把相应概率相加. 其分布函数为

$$F_X(x) = \sum_{x_k \leq x} p_k,$$

因此图像呈右连续的阶梯形, 在 x_k 处的跳跃量恰为 p_k .

例题 2.10 (抽球最大号码)

袋中有编号 1, 2, 3, 4, 5 的五只球, 同时任取三只, 令 X 为三只球中的最大号码. 求 X 的分布律及 $F_X(4.5)$.

先找取值: 最大号码只能为 3, 4, 5. 再分别计数:

$$P(X = 3) = \frac{\binom{3}{3}}{\binom{5}{3}} = \frac{1}{10},$$

$$P(X = 4) = \frac{\binom{3}{2}}{\binom{5}{3}} = \frac{3}{10},$$

$$P(X = 5) = \frac{\binom{4}{2}}{\binom{5}{3}} = \frac{3}{5}.$$

所以

$$F_X(x) = \begin{cases} 0, & x < 3, \\ \frac{1}{10}, & 3 \leq x < 4, \\ \frac{2}{5}, & 4 \leq x < 5, \\ 1, & x \geq 5. \end{cases} \quad F_X(4.5) = \frac{2}{5}.$$

检查: 三项概率之和为 1; 分布函数从 0 逐级累加到 1.

2.2.2 二项分布与 0-1 分布

定义 2.11 (二项分布)

进行 n 次相互独立、条件相同的伯努利试验, 每次事件 A 发生的概率为 p . 令 X 为 A 发生的次数, 则

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n,$$

记作 $X \sim B(n, p)$.

注: 识别二项分布时检查三项: 有固定的 n 次独立重复试验; 每次只关心同一事件是否发生且概率不变; 随机变量统计该事件发生的次数. 少一项都不能直接套二项分布.

例题 2.12 (宿舍考试与“至少一次成功”)

1. 三名学生独立参加考试, 每人通过率为 $2/3$. 求至少两人挂科的概率.
2. 每次独立试验成功概率为 0.5 . 至少进行多少次, 才能使至少成功一次的概率不小于 0.9 ?

第 1 题: 令 X 为挂科人数, 则 $X \sim B(3, 1/3)$. 因此

$$P(X \geq 2) = \binom{3}{2} \left(\frac{1}{3}\right)^2 \frac{2}{3} + \left(\frac{1}{3}\right)^3 = \frac{7}{27}.$$

第 2 题: 令 X 为 n 次试验中的成功次数, 则 $X \sim B(n, 0.5)$. 正面累加项太多, 取反更快:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.5^n \geq 0.9.$$

由 $0.5^n \leq 0.1$ 得最小整数 $n = 4$. 检查: $n = 3$ 时概率为 $0.875 < 0.9$, $n = 4$ 时为 0.9375 .

例题 2.13 (两阶段质检后再用二项分布)

一台仪器以概率 0.7 直接出厂; 否则进入调试, 调试后以概率 0.8 出厂. 各台仪器相互独立. 生产 n 台, 求全部出厂以及恰有两台不能出厂的概率.

先解决单台出厂概率. 令 B 表示“直接出厂”, A 表示“最终出厂”, 则

$$P(A) = P(B) + P(\bar{B})P(A | \bar{B}) = 0.7 + 0.3 \times 0.8 = 0.94.$$

令 X 为最终出厂台数, 则 $X \sim B(n, 0.94)$. 所以

$$P(X = n) = 0.94^n,$$

$$P(X = n - 2) = \binom{n}{2} (0.06)^2 (0.94)^{n-2}.$$

关键转折: 二项分布的参数 p 不是题面第一阶段的 0.7 , 而是先用全概率公式算出的最终成功概率 0.94 .

定义 2.14 (0-1 分布)

若 X 只取 0, 1, 且

$$P(X = 1) = p, \quad P(X = 0) = 1 - p,$$

则称 X 服从参数为 p 的 0-1 分布, 也称伯努利分布. 它是 $B(1, p)$.

2.2.3 几何分布与首次发生问题

定义 2.15 (几何分布)

独立重复进行伯努利试验, 每次成功概率为 p . 令 X 为首次成功时所进行的试验次数, 则

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots,$$

记作 $X \sim Ge(p)$.

例题 2.16 (首次命中)

独立射击同一目标, 每次命中概率为 $1/4$, 直到命中为止. 令 X 为射击次数, 则

$$X \sim Ge\left(\frac{1}{4}\right), \quad P(X = 3) = \left(\frac{3}{4}\right)^2 \frac{1}{4} = \frac{9}{64}.$$

“前 $k - 1$ 次失败, 第 k 次成功”是几何分布公式的直接含义.

例题 2.17 (第四次射击恰为第二次命中)

每次命中概率为 p . 第四次射击恰好是第二次命中, 意味着第四次必命中, 前三次恰有一次命中. 因此

$$P = \binom{3}{1} p(1 - p)^2 \cdot p = 3p^2(1 - p)^2.$$

这不是几何分布, 因为研究的不是“首次命中”; 前三次部分使用二项计数.

注: 考研专项: 若研究第 r 次成功发生时的试验次数, 会得到负二项分布. 普通课程基础阶段只会像上例一样把“最后一次固定成功”和“前面恰有 $r - 1$ 次成功”拆开, 不额外背新公式.

2.2.4 泊松分布与泊松近似

定义 2.18 (泊松分布)

若

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

则称 X 服从参数为 λ 的泊松分布, 记作 $X \sim P(\lambda)$.

泊松分布常描述固定时间或区域内稀有事件发生的次数, 如客流量、故障数或缺陷数.

例题 2.19 (由相邻概率确定参数)

若 $X \sim P(\lambda)$ 且 $P(X = 2) = P(X = 3)$, 则

$$\frac{\lambda^2 e^{-\lambda}}{2!} = \frac{\lambda^3 e^{-\lambda}}{3!},$$

由 $\lambda > 0$ 得 $\lambda = 3$. 于是

$$P(X = 1) = 3e^{-3}.$$

例题 2.20 (从一般分布律识别泊松分布)

设

$$P(X = k) = \frac{C}{k!}, \quad k = 0, 1, 2, \dots$$

求 C 与 $P(X \geq 2)$.

先用分布律的归一性:

$$1 = \sum_{k=0}^{\infty} \frac{C}{k!} = Ce, \quad C = e^{-1}.$$

因此 $X \sim P(1)$. 取反比从 $k = 2$ 向无穷求和更简单:

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{2}{e}.$$

检查: 结果位于 $[0, 1]$ 内.

定理 2.21 (泊松近似二项分布)

若 $X_n \sim B(n, p_n)$, 且当 $n \rightarrow \infty$ 时

$$np_n \rightarrow \lambda > 0,$$

则对固定的非负整数 k ,

$$P(X_n = k) \rightarrow \frac{\lambda^k e^{-\lambda}}{k!}.$$

注: 证明思路 (B/C 级): 把二项概率中的组合数、 p_n^k 和 $(1 - p_n)^{n-k}$ 分别整理, 核心极限是

$$\left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}.$$

泊松定理的严格极限证明属于 C 级, 本阶段不要求复现. 做题时只有题目明确要求近似, 或课程给出适用标准时才用; 否则优先保留精确二项式. 常见经验条件 n 较大、 p 较小只是近似准则, 不是定理条件的替代品.

例题 2.22 (次品数的泊松近似)

独立抽取 100 件产品, 每件为次品的概率为 0.05. 令 X 为次品数, 则

$$X \sim B(100, 0.05) \approx P(5).$$

若题目要求使用泊松近似,

$$P(X \geq 3) \approx 1 - e^{-5} \left(1 + 5 + \frac{5^2}{2!}\right).$$

2.2.5 超几何分布: 无放回抽样

定义 2.23 (超几何分布)

总体有 N 件, 其中 M 件为目标类别; 无放回抽取 n 件, 令 X 为抽到的目标件数, 则

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}},$$

其中

$$\max\{0, n - (N - M)\} \leq k \leq \min\{n, M\}.$$

记作 $X \sim H(N, M, n)$ (符号顺序以教材为准) .

注: 二项分布与超几何分布的关键区别:

比较项	二项分布	超几何分布
抽样方式	有放回或每次概率不变	无放回
各次结果	独立	一般不独立
成功概率	每次均为 p	随抽样结果改变

当总体很大、抽样比例很小时, 无放回对总体比例的影响很小, 此时可近似使用

$$B\left(n, \frac{M}{N}\right).$$

是否使用近似应服从题目要求.

例题 2.24 (抽到第一件合格品所需次数)

一批产品有 10 件合格品、3 件次品, 逐件抽取, 直到抽到合格品为止. 令 X 为抽取次数, 分别讨论有放回与无放回.

有放回: 每次抽到合格品的概率保持 10/13, 故

$$X \sim Ge\left(\frac{10}{13}\right), \quad P(X = k) = \left(\frac{3}{13}\right)^{k-1} \frac{10}{13}.$$

无放回: 最多连续抽出三件次品, 所以 $X \in \{1, 2, 3, 4\}$. 依次使用乘法公式:

$$\begin{aligned} P(X = 1) &= \frac{10}{13}, \\ P(X = 2) &= \frac{3}{13} \frac{10}{12}, \\ P(X = 3) &= \frac{3}{13} \frac{2}{12} \frac{10}{11}, \\ P(X = 4) &= \frac{3}{13} \frac{2}{12} \frac{1}{11}. \end{aligned}$$

检查: 四项之和为 1. 本题也说明“第 k 次首次抽到合格品”比“第 k 次抽到合格品”条件更强, 不能误用抽签原理.

小结

离散型随机变量的分布由“可能取值 + 对应概率”组成, 计算概率就是对互斥取值求和. 识别常见分布时先读背景: 固定次数的成功数用二项分布, 首次成功次数用几何分布, 固定区间内的稀有计数常用泊松分布, 无放回抽样用超几何分布; 使用前必须核对独立性、是否放回、参数含义和取值范围.

2.3 连续型随机变量与概率密度

2.3.1 密度的含义与基本条件

定义 2.25 (连续型随机变量与概率密度)

若存在非负可积函数 f_X , 使随机变量 X 的分布函数可写为

$$F_X(x) = \int_{-\infty}^x f_X(t) dt,$$

则称 X 为连续型随机变量, 称 f_X 为 X 的概率密度函数, 简称密度.

密度 $f_X(x)$ 不是点概率. 它的高低反映 X 落在 x 附近的概率集中程度:

$$P(x < X \leq x + \Delta x) = \int_x^{x+\Delta x} f_X(t) dt \approx f_X(x)\Delta x.$$

改变密度在有限个点处的函数值不会改变任何积分, 也不会改变随机变量的分布.

性质 2.26 (密度的充要条件)

函数 f 能成为某个连续型随机变量的密度, 当且仅当

$$f(x) \geq 0, \quad \int_{-\infty}^{+\infty} f(x) dx = 1.$$

定理 2.27 (密度、分布函数与区间概率)

若 X 的密度为 f_X , 则

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad P(a < X \leq b) = \int_a^b f_X(x) dx.$$

在 f_X 的连续点处,

$$F'_X(x) = f_X(x).$$

证明: **证明目标:** 从密度定义推出区间概率公式, 并说明何时能由分布函数求密度.

出发点: 密度定义、分布函数的区间概率公式和微积分基本定理.

关键转换: 把两个变上限积分相减, 化为固定区间积分.

正式推导:

$$\begin{aligned} P(a < X \leq b) &= F_X(b) - F_X(a) \\ &= \int_{-\infty}^b f_X(t) dt - \int_{-\infty}^a f_X(t) dt \\ &= \int_a^b f_X(t) dt. \end{aligned}$$

若 f_X 在 x 处连续, 则由微积分基本定理,

$$\frac{d}{dx} \int_{-\infty}^x f_X(t) dt = f_X(x).$$

证明小结: “在哪个区间求概率, 就在哪个区间对密度积分”. 由 F_X 求 f_X 时只需在各连续区间求导; 有限个分段点处的密度值可任取, 但不能把分布函数的跳跃部分误当成普通导数. ■

定理 2.28 (连续型随机变量的点概率)

若 X 为连续型随机变量, 则对任意 $a \in \mathbb{R}$,

$$P(X = a) = 0.$$

因此区间端点是否包含不影响概率, 例如

$$P(a < X < b) = P(a \leq X \leq b) = \int_a^b f_X(x) dx.$$

注: 证明思路 (B 级): 单点集合的积分为零,

$$P(X = a) = \int_{\{a\}} f_X(x) dx = 0;$$

也可用连续型分布函数连续, 得到 $F_X(a) - F_X(a - 0) = 0$.

“点概率为零”不表示该数值逻辑上不可能出现; 它表示单个点在连续尺度上没有正的概率质量. 不能把这一结论用于离散型随机变量.

注: [三个容易混淆的“连续”]

对象	正确结论	常见误判
连续型随机变量 X	其分布可由某个密度积分表示	不是说 X 是普通连续函数
分布函数 F_X	连续型情形下一定连续	分布函数连续未必在更高阶理论中推出 X 连续型
概率密度 f_X	只要求非负、可积且积分为 1	密度本身不必连续

例题 2.29 (判断随机变量的类型)

下列对象分别属于哪一类随机变量?

1. 已知 X 具有概率密度;
2. 令 $U = 1$ 表示事件 $\{X > 2\}$ 发生, 否则 $U = 0$;
3. 相邻两次故障之间的时间, 以及一段固定时间内的故障次数;
4. 分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ x/2, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

为什么这样判断: 先看取值是否可数, 再看分布函数是否连续. 第 1 个是连续型; 第 2 个只取 0, 1, 是离散型; 故障间隔通常按连续型建模, 而故障次数只能取非负整数, 是离散型. 第 4 个在区间内连续增长、又在 $x = 1$ 处跳跃, 因此既不是纯离散型, 也不是连续型, 而是混合型.

结果检查: “原变量连续”不能推出“它的任意函数仍连续”, 第 2 个对象正是反例.

2.3.2 从密度求参数、概率与分布函数

注: [连续型综合题的四步路线] 面对分段密度时, 不要一上来就积分. 建议固定按以下顺序处理:

1. **查合法性:** 先检查非负性, 再用 $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ 确定未知参数;
2. **画支持集:** 标出密度非零的区间, 目标事件只与支持集的交集有关;
3. **按分点积分:** 求概率时在目标区间积分, 求分布函数时把 x 经过的每个分点都列出来;
4. **做结果检查:** 概率应落在 $[0, 1]$, 分布函数应单调、连续且两端极限分别为 0, 1.

由已有密度构造新密度时必须重查归一性. 若 f_1, f_2 都是概率密度, 则 $af_1 + bf_2$ 在 $a, b \geq 0$ 且 $a + b = 1$ 时仍是密度; $f(1-x)$ 经变量代换后总积分不变. 相反, $2f(x)$ 的总积分为 2, 而 $f_1(x)f_2(x)$ 的总积分也未必等于 1, 一般不能直接作为密度.

例题 2.30 (分段密度综合题)

设

$$f_X(x) = \begin{cases} x, & 0 < x < 1, \\ k(2-x), & 1 < x < 2, \\ 0, & \text{其他.} \end{cases}$$

求 k 、 $P\left(\frac{1}{2} < X \leq \frac{3}{2}\right)$ 、 $F_X(x)$ 以及 $P\left(X > \frac{1}{2} \mid X \leq 1\right)$.

第 1 步: 用归一性确定参数.

$$1 = \int_0^1 x dx + \int_1^2 k(2-x) dx = \frac{1}{2} + \frac{k}{2},$$

故 $k = 1$.

第 2 步: 在目标区间积分.

$$\begin{aligned} P\left(\frac{1}{2} < X \leq \frac{3}{2}\right) &= \int_{1/2}^1 x dx + \int_1^{3/2} (2-x) dx \\ &= \frac{3}{8} + \frac{3}{8} = \frac{3}{4}. \end{aligned}$$

第 3 步: 先用支持集秒杀首尾两段, 再计算中间.

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{x^2}{2}, & 0 \leq x < 1, \\ 1 - \frac{(2-x)^2}{2}, & 1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

第 4 步: 条件概率.

$$\begin{aligned} P\left(X > \frac{1}{2} \mid X \leq 1\right) &= \frac{P(1/2 < X \leq 1)}{P(X \leq 1)} \\ &= \frac{F_X(1) - F_X(1/2)}{F_X(1)} = \frac{1/2 - 1/8}{1/2} = \frac{3}{4}. \end{aligned}$$

检查: F_X 连续、单调不减, 且首尾极限为 0, 1.

例题 2.31 (全实轴上的双指数密度)

设

$$f_X(x) = Ae^{-|x|}, \quad -\infty < x < +\infty.$$

求 A 、 $P(0 < X < 1)$ 和 $F_X(x)$.

由偶函数与归一性,

$$1 = 2A \int_0^{\infty} e^{-x} dx = 2A, \quad A = \frac{1}{2}.$$

所以

$$P(0 < X < 1) = \frac{1}{2} \int_0^1 e^{-x} dx = \frac{1}{2}(1 - e^{-1}).$$

分布函数需按 $x = 0$ 分段:

$$F_X(x) = \begin{cases} \frac{1}{2}e^x, & x < 0, \\ 1 - \frac{1}{2}e^{-x}, & x \geq 0. \end{cases}$$

本例的支持集没有有限上下界, 因此 $F_X(x)$ 在有限点处既不取 0, 也不取 1, 只在两端趋近.

例题 2.32 (验证给定函数能否作为密度)

判断下列函数能否作为概率密度:

$$f_1(x) = \begin{cases} \frac{1}{2}, & |x| < 2, \\ 0, & \text{其他}, \end{cases} \quad f_2(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

验证目标: 检查非负性和归一性.

出发点: 密度的充要条件.

关键转换: 只在函数非零的支持区间积分.

逐项验证: 两者都非负, 但

$$\int_{-\infty}^{\infty} f_1(x) dx = \int_{-2}^2 \frac{1}{2} dx = 2 \neq 1,$$

所以 f_1 不是密度; 而

$$\int_{-\infty}^{\infty} f_2(x) dx = \int_0^{\infty} e^{-x} dx = 1,$$

所以 f_2 是密度.

验证小结: 图像“看起来像密度”不够; 高度必须与支持区间长度共同满足总面积为 1.

例题 2.33 (灯管寿命与二项分布的衔接)

灯管寿命 X 的密度为

$$f_X(x) = \begin{cases} \frac{1000}{x^2}, & x > 1000, \\ 0, & \text{其他}. \end{cases}$$

则

$$P(X > 1500) = \int_{1500}^{\infty} \frac{1000}{x^2} dx = \frac{2}{3}.$$

任取 5 支且寿命相互独立, 令 Y 为寿命超过 1500 小时的支数, 则

$$Y \sim B\left(5, \frac{2}{3}\right),$$

$$P(Y \geq 2) = 1 - \left(\frac{1}{3}\right)^5 - \binom{5}{1} \frac{2}{3} \left(\frac{1}{3}\right)^4 = \frac{232}{243}.$$

方法选择: 先用连续型积分求单支成功概率, 再用二项分布统计五支中的成功数.

小结

连续型随机变量用密度描述分布: 密度非负且总积分为 1, 区间概率由积分得到, 单点概率为零. 做综合题时依次完成“归一化求参数—按支持集分段积分—由积分写分布函数—必要时求导还原密度”; 遇到第二层计数问题, 再把单次概率交给二项分布.

2.4 三种常见连续型分布

2.4.1 均匀分布

定义 2.34 (均匀分布)

若

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{其他}, \end{cases}$$

则称 X 在区间 (a, b) 上服从均匀分布, 记作 $X \sim U(a, b)$. 其分布函数为

$$F_X(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases}$$

均匀分布的背景是“在区间内按长度等可能”. 因此对 $(c, d) \subseteq (a, b)$,

$$P(c < X < d) = \frac{d-c}{b-a}.$$

例题 2.35 (均匀分布与方程实根)

设 $X \sim U(1, 6)$. 求方程

$$t^2 + Xt + 1 = 0$$

有实根的概率.

方程有实根当且仅当判别式非负:

$$X^2 - 4 \geq 0.$$

结合 $X \in (1, 6)$, 事件化简为 $\{X \geq 2\}$. 均匀分布按长度计算,

$$P(X \geq 2) = \frac{6-2}{6-1} = \frac{4}{5}.$$

关键转折: 解出不等式后必须与随机变量的支持集取交集.

例题 2.36 (由随机取点识别均匀分布)

在区间 $(0, 2)$ 上随机取一点, 令 X 为落点位置. “任意小区间的概率与长度成正比”正是均匀分布背景, 因此

$$X \sim U(0, 2), \quad f_X(x) = \begin{cases} \frac{1}{2}, & 0 < x < 2, \\ 0, & \text{其他.} \end{cases}$$

2.4.2 指数分布与无记忆性

定义 2.37 (指数分布)

若 $\lambda > 0$ 且

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

则称 X 服从参数为 λ 的指数分布, 记作 $X \sim E(\lambda)$. 其分布函数为

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

指数分布常用于无明显老化阶段的等待时间或寿命. 对 $a > 0$,

$$P(X > a) = e^{-\lambda a}.$$

若教材用尺度参数 $\theta = 1/\lambda$, 所有公式中的 λ 应替换为 $1/\theta$.

定理 2.38 (指数分布的无记忆性)

若 $X \sim E(\lambda)$, 则对 $s, t > 0$,

$$P(X > s + t \mid X > s) = P(X > t).$$

证明: **证明目标:** 证明已经等待 s 后, 再等待超过 t 的概率与从零开始等待超过 t 的概率相同.

出发点: 条件概率定义与指数分布的尾概率公式.

关键转换: 因为 $s + t > s$,

$$\{X > s + t\} \subseteq \{X > s\}.$$

正式推导:

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} = P(X > t). \end{aligned}$$

证明小结: 无记忆性依赖指数尾概率的乘法结构. 它不是所有寿命分布的通性, 也不表示现实设备永不老化. ■

例题 2.39 (指数分布综合题)

设 $X \sim E(2)$, 求密度、分布函数、

$$P(X \leq a + 1 \mid X > a) \quad (a > 0),$$

$F_X(2)$ 和 $P(X = 1)$.

$$f_X(x) = \begin{cases} 2e^{-2x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-2x}, & x \geq 0. \end{cases}$$

利用条件事件中的对立关系和无记忆性,

$$\begin{aligned} P(X \leq a+1 \mid X > a) &= 1 - P(X > a+1 \mid X > a) \\ &= 1 - P(X > 1) = 1 - e^{-2}. \end{aligned}$$

此外

$$F_X(2) = 1 - e^{-4}, \quad P(X = 1) = 0.$$

2.4.3 正态分布与标准化

定义 2.40 (正态分布)

若 $\sigma > 0$ 且

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < +\infty,$$

则称 X 服从均值参数为 μ 、方差参数为 σ^2 的正态分布, 记作

$$X \sim N(\mu, \sigma^2).$$

密度曲线关于 $x = \mu$ 对称, 在 $x = \mu$ 处最高; 越远离 μ , 附近概率密度越小. 因此

$$P(X \leq \mu) = P(X \geq \mu) = \frac{1}{2}.$$

性质 2.41 (3σ 原则)

若 $X \sim N(\mu, \sigma^2)$, 则近似有

$$P(|X - \mu| \leq \sigma) \approx 0.6827,$$

$$P(|X - \mu| \leq 2\sigma) \approx 0.9545,$$

$$P(|X - \mu| \leq 3\sigma) \approx 0.9973.$$

超出 $\mu \pm 3\sigma$ 的结果很少见, 但并非逻辑上不可能.

定义 2.42 (标准正态分布)

当 $\mu = 0, \sigma = 1$ 时, 称 Z 服从标准正态分布, 记作 $Z \sim N(0, 1)$. 其密度与分布函数记为

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = P(Z \leq z).$$

定理 2.43 (正态标准化)

若 $X \sim N(\mu, \sigma^2)$, 则

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

因此

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

注: 正态标准化的分布函数证明属于 C 级: 本阶段不要求复现换元积分的完整细节, 只需掌握“减均值、除标准差”, 并记住参数写的是 σ^2 、标准化分母写的是 σ .

由标准正态密度的对称性,

$$\Phi(-a) = 1 - \Phi(a), \quad \Phi(0) = \frac{1}{2},$$

且对 $a > 0$,

$$P(|Z| \leq a) = 2\Phi(a) - 1.$$

例题 2.44 (只用对称性, 不必标准化)

若 $X \sim N(2, \sigma^2)$ 且 $P(2 < X < 4) = 0.3$, 则区间 $(0, 2)$ 与 $(2, 4)$ 关于 $\mu = 2$ 对称, 所以

$$P(0 < X < 2) = 0.3.$$

同理, 若 $P(X \leq 4) = 0.5$, 则正态密度的对称轴必须为 4, 故 $\mu = 4$.

例题 2.45 (一次函数仍服从正态)

若 $X \sim N(\mu, \sigma^2)$, 则

$$aX + b \sim N(a\mu + b, a^2\sigma^2), \quad a \neq 0.$$

例如 $X \sim N(1, 1)$ 时, 要使 $X + C \sim N(0, 1)$, 必须取 $C = -1$.

例题 2.46 (公交车门高度)

成年男子身高 $X \sim N(165, 36)$ (单位: 厘米). 若车门碰头率不超过 1%, 且已知 $\Phi(2.33) = 0.99$, 求车门高度的最低设计值.

设车门高度为 h . 碰头率不超过 1% 等价于

$$P(X \leq h) \geq 0.99.$$

标准化:

$$\Phi\left(\frac{h - 165}{6}\right) \geq \Phi(2.33).$$

由于 Φ 严格递增,

$$\frac{h - 165}{6} \geq 2.33, \quad h \geq 178.98.$$

实际设计可取 179 厘米. 检查: 这里的 36 是方差, 所以标准差为 6.

例题 2.47 (由中心区间概率比较标准差)

设

$$X \sim N(\mu_1, \sigma_1^2), \quad Y \sim N(\mu_2, \sigma_2^2),$$

且

$$P(|X - \mu_1| < 1) > P(|Y - \mu_2| < 1).$$

标准化后

$$2\Phi(1/\sigma_1) - 1 > 2\Phi(1/\sigma_2) - 1.$$

因 Φ 严格递增, 得 $\sigma_1 < \sigma_2$. 标准差越小, 分布越集中.

例题 2.48 (从二次指数识别正态分布)

已知某密度在全实轴上形如

$$f_X(x) = A \exp\left(-\frac{1}{4}x^2 + x\right).$$

识别 X 的正态参数, 并求

$$P(-2\sqrt{2} + 2 < X < 2\sqrt{2} + 2)$$

(用 Φ 表示).

先配方:

$$-\frac{1}{4}x^2 + x = -\frac{(x-2)^2}{4} + 1.$$

与

$$-\frac{(x-\mu)^2}{2\sigma^2}$$

比较, 得 $\mu = 2$, $\sigma^2 = 2$. 因此

$$\frac{X-2}{\sqrt{2}} \sim N(0, 1),$$

$$P(-2\sqrt{2} + 2 < X < 2\sqrt{2} + 2) = P(-2 < Z < 2) = 2\Phi(2) - 1.$$

题目若不要求 A , 无需先计算归一化常数.

小结

均匀分布按区间长度计算, 指数分布用尾概率 $e^{-\lambda a}$ 并具有无记忆性, 正态分布先找对称轴, 再决定是否标准化. 使用时必须核对参数口径: 均匀分布看区间端点, 指数分布区分速率 λ 与尺度 θ , 正态分布区分方差 σ^2 与标准差 σ .

2.5 随机变量函数的分布

2.5.1 离散型: 合并产生同一函数值的概率

若离散型随机变量 X 的分布律为 $P(X = x_k) = p_k$, 令 $Y = g(X)$, 则:

1. 先计算所有 $g(x_k)$, 去掉重复值, 得到 Y 的全部可能取值;
2. 对每个 y , 把所有满足 $g(x_k) = y$ 的概率相加:

$$P(Y = y) = \sum_{k: g(x_k)=y} p_k.$$

例题 2.49 (离散型函数的分布)

已知

X	-1	0	1	2
P	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{4}$

令 $Y = X^2$, 求 Y 的分布律.

由 $Y = X^2$, 可能取值为 0, 1, 4. 其中 $X = -1$ 与 $X = 1$ 都映射到 $Y = 1$, 所以

$$P(Y = 0) = P(X = 0) = \frac{1}{3},$$

$$P(Y = 1) = P(X = -1) + P(X = 1) = \frac{1}{4} + \frac{1}{6} = \frac{5}{12},$$

$$P(Y = 4) = P(X = 2) = \frac{1}{4}.$$

检查: 三项概率之和为 1. 易错点是只取 $X = 1$ 而漏掉 $X = -1$.

2.5.2 连续型: 分布函数法

定理 2.50 (一维随机变量函数的分布函数法)

设连续型随机变量 X 的分布已知, $Y = g(X)$. 对任意实数 y ,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y).$$

把不等式 $g(X) \leq y$ 化为关于 X 的取值集合, 再用 X 的分布函数或密度计算, 即可得到 F_Y ; 若 F_Y 在相应区间可导, 再由

$$f_Y(y) = F'_Y(y)$$

求密度.

注: 分布函数法是通法, 尤其适合 g 非单调、多对一或含绝对值的情形. 固定流程为:

先定 Y 的范围 $\longrightarrow$ 写 $P(g(X) \leq y)$ $\longrightarrow$ 化成 X 的范围 $\longrightarrow$ 求概率并求导.

若 g 在 X 的支持区间上严格单调且可导, 也可使用反函数密度公式, 但本章优先掌握不易漏解的分布函数法.

例题 2.51 (单调线性变换)

设

$$f_X(x) = \begin{cases} \frac{x}{8}, & 0 < x < 4, \\ 0, & \text{其他}, \end{cases} \quad Y = 2X + 8.$$

求 Y 的密度.

推导目标: 用分布函数法求 Y 的分布.

出发点: $X \in (0, 4)$, 故 $Y \in (8, 16)$.

关键转换: 对中间区间 $8 \leq y < 16$,

$$\{Y \leq y\} = \{2X + 8 \leq y\} = \left\{X \leq \frac{y-8}{2}\right\}.$$

正式推导:

$$F_Y(y) = \begin{cases} 0, & y < 8, \\ \int_0^{(y-8)/2} \frac{x}{8} dx = \frac{(y-8)^2}{64}, & 8 \leq y < 16, \\ 1, & y \geq 16. \end{cases}$$

于是

$$f_Y(y) = \begin{cases} \frac{y-8}{32}, & 8 < y < 16, \\ 0, & \text{其他.} \end{cases}$$

推导小结: 先确定 Y 的支持集可以直接写出 0 与 1 两段, 也可防止求导后漏写密度的取值范围.

例题 2.52 (绝对值变换: 两侧概率要相加)

设

$$f_X(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < +\infty,$$

令 $Y = |X|$. 求 Y 的密度.

因为 $Y \geq 0$,

$$F_Y(y) = 0, \quad y < 0.$$

当 $y \geq 0$ 时,

$$\begin{aligned} F_Y(y) &= P(|X| \leq y) = P(-y \leq X \leq y) \\ &= \int_{-y}^y \frac{1}{2}e^{-|x|} dx = 1 - e^{-y}. \end{aligned}$$

所以

$$f_Y(y) = \begin{cases} e^{-y}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

结果是参数为 1 的指数分布. 检查: 密度非负且在 $(0, \infty)$ 上积分为 1.

例题 2.53 (均匀分布经过绝对值变换)

设 $X \sim U(-1, 2)$, 令 $Y = |X|$, 求 Y 的密度.

先画 $y = |x|$ 在 $x \in [-1, 2]$ 上的图像, 可见 $Y \in [0, 2]$. 水平线在 $0 < y < 1$ 时与左右两支相交, 在 $1 < y < 2$ 时只与右支相交, 因此必须以 $y = 1$ 分段.

当 $0 \leq y < 1$ 时,

$$F_Y(y) = P(-y \leq X \leq y) = \frac{2y}{3}.$$

当 $1 \leq y < 2$ 时, 左端已超出 X 的支持集,

$$F_Y(y) = P(-1 \leq X \leq y) = \frac{y+1}{3}.$$

故

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ \frac{2y}{3}, & 0 \leq y < 1, \\ \frac{y+1}{3}, & 1 \leq y < 2, \\ 1, & y \geq 2, \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{2}{3}, & 0 < y < 1, \\ \frac{1}{3}, & 1 < y < 2, \\ 0, & \text{其他.} \end{cases}$$

关键转折: 非单调变换可能使不同的 x 映射到同一个 y , 且交点个数变化的位置就是新的分段点.

例题 2.54 (正态绝对偏差)

若 $X \sim N(1, 4)$, 令 $Y = |X - 1|$. 当 $y \geq 0$ 时,

$$\begin{aligned} F_Y(y) &= P(|X - 1| \leq y) \\ &= P\left(\left|\frac{X - 1}{2}\right| \leq \frac{y}{2}\right) = 2\Phi\left(\frac{y}{2}\right) - 1. \end{aligned}$$

因此

$$f_Y(y) = \begin{cases} \varphi(y/2), & y > 0, \\ 0, & y < 0. \end{cases}$$

这里利用的是正态标准化, 而不是直接积分正态密度.

2.5.3 概率积分变换

定理 2.55 (概率积分变换)

若连续型随机变量 X 的分布函数 F_X 在其有效取值区间上严格递增, 则

$$Y = F_X(X) \sim U(0, 1).$$

注: 证明思路 (B 级): 当 $0 < y < 1$ 时, 利用严格递增性和反函数,

$$\begin{aligned} F_Y(y) &= P(F_X(X) \leq y) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

再补上 $y < 0$ 时为 0、 $y \geq 1$ 时为 1, 即得到 $U(0, 1)$ 的分布函数. 若 F_X 不严格递增, 可用广义逆作更一般的表述; 本阶段不要求展开.

注: 考研专项: 概率积分变换的用途. 它把一般连续分布转成均匀分布, 是随机模拟和蒙特卡洛方法的基础之一. 基础阶段会识别并使用结论即可.

小结

离散型函数分布通过“代入取值、合并同像概率”得到; 连续型函数分布优先使用分布函数法. 使用时先确定 Y 的支持集, 再把 $Y \leq y$ 转为关于 X 的事件. 函数非单调时必须检查一个 y 对应几个 x , 交点个数发生变化的位置就是分段点.

2.6 本章总结

2.6.1 知识结构

$$\begin{aligned}
 &\text{随机变量 } X \longrightarrow F_X(x) = P(X \leq x) \\
 &\longrightarrow \begin{cases} \text{离散型: } P(X = x_k) = p_k, & F_X(x) = \sum_{x_k \leq x} p_k, \\ \text{连续型: } F_X(x) = \int_{-\infty}^x f_X(t) dt, \end{cases} \\
 &\longrightarrow Y = g(X) \text{ 的分布.}
 \end{aligned}$$

2.6.2 核心公式与适用条件

公式	内容	使用前检查
区间概率	$P(a < X \leq b) = F_X(b) - F_X(a)$	端点开闭、左极限
离散型概率	$P(X \in A) = \sum_{x_k \in A} p_k$	取值是否列全
连续型概率	$P(a < X < b) = \int_a^b f_X(x) dx$	密度与支持集
二项分布	$\binom{n}{k} p^k (1-p)^{n-k}$	固定次数、独立、概率不变
几何分布	$(1-p)^{k-1} p$	首次成功、第 k 次停止
泊松分布	$\lambda^k e^{-\lambda} / k!$	$k = 0, 1, \dots$
超几何分布	$\binom{M}{k} \binom{N-M}{n-k} / \binom{N}{n}$	无放回、取值上下界
指数尾概率	$P(X > a) = e^{-\lambda a}$	$X \sim E(\lambda), a > 0$
正态标准化	$(X - \mu) / \sigma \sim N(0, 1)$	分母是标准差 σ
函数分布	$F_Y(y) = P(g(X) \leq y)$	先求 Y 的支持集

2.6.3 常见题型与方法选择

- 判断函数是否合法:** 分布函数查单调不减、右连续、两端极限; 分布律查非负与求和为 1; 密度查非负与积分为 1.
- 由分布函数算概率:** 把目标区间写成两个左侧事件之差, 离散情形保留左极限.
- 离散型求分布律:** 先列全部取值, 再逐个求概率; 复杂的最后一项可用总概率为 1 补出.

4. **识别常见分布:** 先读试验结构, 不要只凭公式外形; 特别区分二项、几何和超几何.
5. **连续型综合题:** 归一化求参数, 再按支持集积分; 分布函数先写支持集外的 0、1 两段.
6. **正态题:** 先看对称性能否直接解决, 再标准化; 看到方差参数先开方得到标准差.
7. **随机变量函数:** 离散型合并同像概率; 连续型用分布函数法, 非单调时画 $y = g(x)$ 并找分段点.

2.6.4 重点辨析

概念	本质区别	常见误判
分布函数与密度	F_X 是累计概率, f_X 是概率集中程度	把 $f_X(a)$ 当作 $P(X = a)$
离散与连续	前者对取值求和, 后者对区间积分	看到“随机”就默认等可能
互斥与独立	互斥是不同时发生, 独立是不影响概率	把互斥事件当作独立事件
二项与超几何	二项要求概率不变, 超几何是无放回	无放回仍直接套二项
几何与二项	几何统计首次成功所需次数, 二项统计固定次数内成功数	忽略“直到”“为止”
方差与标准差	正态记号写 σ^2 , 标准化除以 σ	直接除以题中的方差数字
点概率为零与不可能	连续型单点概率为零, 但单点可作为结果出现	由 $P(A) = 0$ 推出 $A = \emptyset$

2.6.5 易错点

1. 分布函数定义域是整个实数轴, 不能只写随机变量的支持区间.
2. 分布函数右连续; 离散型跳跃点处等号应放在右侧一段.
3. 分布律和密度都要先写取值范围或支持集, 不能只写公式主体.
4. 泊松近似不能擅自代替精确二项概率.
5. 密度在个别点处的值不影响分布, 但整个区间上的密度不能随意改.
6. 连续型区间端点可忽略, 离散型端点通常不可忽略.
7. 求 $Y = g(X)$ 时, 不能默认 g 单调; 例如平方和绝对值会合并两侧概率.

2.6.6 证明掌握清单

- **A 级:** 会完整验证分布函数、分布律和密度; 会由密度推出分布函数和区间概率; 会处理分布函数跳跃点; 会用分布函数法推导 $Y = g(X)$ 的分布; 会证明指数分布无记忆性.
- **B 级:** 会解释二项、几何、泊松等公式的试验背景; 理解泊松近似的条件与限制; 能说明连续型单点概率为零; 理解概率积分变换的证明骨架.
- **C 级:** 泊松定理严格极限证明、正态标准化的完整积分证明、分布函数左极限的严格分析证明, 以及常见分布归一化的重复代数推导, 本阶段不要求复现.

2.6.7 自测题

练习 2.56 (分布函数与跳跃)

设

$$F(x) = \begin{cases} 0, & x < 0, \\ c, & 0 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

确定 c 的允许范围, 并写出 $P(X = 0)$ 、 $P(X = 2)$ 与 $P(0 < X < 2)$.

练习 2.57 (分布选择)

分别判断下列随机变量更适合二项、几何、泊松还是超几何分布, 并说明条件:

1. 独立检查 20 件产品中的次品数;
2. 无放回抽取 5 件产品中的次品数;
3. 直到第一次出现次品所检查的件数;
4. 一小时内到达服务台的人数.

练习 2.58 (密度与分布函数)

设

$$f(x) = \begin{cases} kx, & 0 < x < 2, \\ 0, & \text{其他.} \end{cases}$$

求 k 、分布函数 F 、 $P(X > 1)$, 并验证所得 F 满足分布函数条件.

练习 2.59 (正态标准化)

设 $X \sim N(10, 4)$, 用 Φ 表示

$$P(8 < X < 13), \quad P(|X - 10| \leq 3).$$

练习 2.60 (随机变量函数)

设 $X \sim U(-2, 1)$, 令 $Y = X^2$. 先画 $y = x^2$ 在支持区间上的图像, 再求 Y 的分布函数和密度. 检查分段点为何不是只有 0.

注: [本章自检] 完成本章后, 应能回答:

1. 为什么分布函数能表示随机变量的全部概率信息?
2. 如何从题目语言判断该求和还是积分?
3. 五种离散分布和三种连续分布各自的参数、支持集与典型背景是什么?
4. 为什么连续型单点概率为零, 却不能简单理解成“绝不可能取到”?
5. 面对 $Y = g(X)$, 为什么分布函数法比死背变换公式更稳妥?

第三章 多维随机变量及其分布

本章导读

第二章只用一个数描述一次试验结果, 但现实问题常由多个指标共同决定. 例如一件产品同时有长度和重量, 一名学生同时有数学和英语成绩. 若只研究其中一个指标, 就会丢失它们如何共同变化的信息. 本章把同一次试验产生的多个随机变量放在一起, 研究它们的**联合分布**、各自的**边缘分布**、已知一个变量后的**条件分布**以及变量之间的**独立性**.

本章的计算主线只有两条:

离散型: 对符合条件的格点概率求和, 连续型: 在符合条件的平面区域上积分.

最后再把 (X, Y) 通过 $Z = g(X, Y)$ 压缩成一个新随机变量. 学习时始终先问: **取值区域是什么?**
目标事件在图上是哪一块?

3.1 二维随机变量与联合分布函数

3.1.1 为什么需要二维随机变量

定义 3.1 (二维随机变量)

在同一个随机试验的样本空间 Ω 上定义两个随机变量 $X = X(\omega)$ 与 $Y = Y(\omega)$, 把它们作为一个整体

$$(X, Y) = (X(\omega), Y(\omega)),$$

称为二维随机变量或随机向量. 一次试验结果 ω 会同时确定平面上的一个随机点

(X, Y) .

注: [“同一个样本空间”是关键] 二维随机变量不是把两个毫无关系的数机械并排, 而是同一次随机试验的两个观测指标. 可以分别研究 X 、 Y , 也可以研究它们同时落入某区域的概率. 推广到 n 个指标, 就得到 n 维随机变量 $(X_1, \dots, X_n)$.

3.1.2 联合分布函数与矩形概率

定义 3.2 (联合分布函数)

二维随机变量 (X, Y) 的联合分布函数定义为

$$F(x, y) = P(X \leq x, Y \leq y), \quad -\infty < x, y < +\infty.$$

几何上, $F(x, y)$ 是随机点落在点 (x, y) 左下方无界矩形中的概率.

定理 3.3 (由联合分布函数求矩形概率)

若 $x_1 < x_2$ 、 $y_1 < y_2$, 则

$$\begin{aligned} P(x_1 < X \leq x_2, y_1 < Y \leq y_2) \\ = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \end{aligned}$$

证明: 证明目标: 从联合分布函数中扣出半开矩形的概率.

出发点: 联合分布函数表示左下方无界矩形的概率.

关键转换: 先取大矩形, 再减去左条和下条; 二者重叠的左下角被减了两次, 需加回一次.

正式推导: 记 $A = \{X \leq x_2, Y \leq y_2\}$, 则利用容斥关系,

$$\begin{aligned} P(x_1 < X \leq x_2, y_1 < Y \leq y_2) &= P(A) - P(X \leq x_1, Y \leq y_2) \\ &\quad - P(X \leq x_2, Y \leq y_1) + P(X \leq x_1, Y \leq y_1) \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \end{aligned}$$

证明小结: 真正起作用的是矩形的包含关系和容斥原理. 四项符号可记为“右上减左上、减右下、加左下”.

定理 3.4 (联合分布函数的基本性质)

任意联合分布函数 F 满足:

1. $0 \leq F(x, y) \leq 1$;
2. 对每个变量分别单调不减, 并分别右连续;
3. $F(-\infty, y) = F(x, -\infty) = 0, F(+\infty, +\infty) = 1$;
4. 对任意 $x_1 < x_2, y_1 < y_2$, 矩形增量

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0.$$

最后一条称为二重单调性; 仅分别单调还不够.

例题 3.5 (分别单调不等于合法联合分布函数)

判断某个二元函数是否能作联合分布函数时, 即使它取值在 $[0, 1]$ 、分别单调且满足端点极限, 也必须检查任意矩形增量非负. 若能找到一个矩形使四项差为负, 就等于得到一个“负概率”, 该函数立即被排除.

方法选择: 这类题不必尝试还原随机变量, 只需逐条检查充要条件.

3.1.3 边缘分布函数与独立性

定义 3.6 (边缘分布函数)

忽略另一个变量后, X 与 Y 的分布函数称为边缘分布函数:

$$F_X(x) = F(x, +\infty), \quad F_Y(y) = F(+\infty, y).$$

“边缘”就是高维分布中某个低维分量自己的分布.

定义 3.7 (随机变量的独立性)

若对任意实数 x, y 都有

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

则称 X 与 Y 相互独立.

例题 3.8 (由联合分布函数求边缘并判断独立)

设

$$F(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-\lambda y}), & x \geq 0, y \geq 0, \\ 0, & \text{其他,} \end{cases} \quad \lambda > 0.$$

求边缘分布函数并判断 X, Y 是否独立.

令另一个变量趋于 $+\infty$:

$$F_X(x) = \begin{cases} 1 - e^{-x}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad F_Y(y) = \begin{cases} 1 - e^{-\lambda y}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

对所有 x, y , 原联合分布函数正好等于 $F_X(x)F_Y(y)$, 故 X, Y 独立.

检查: 两个边缘分别是参数为 1 和 λ 的指数分布; 联合分布中的乘积结构与独立性一致.

小结

本节把一次试验的两个指标组成随机点 (X, Y) . 联合分布函数记录左下方矩形概率; 求有限矩形概率用四项容斥; 求边缘时让另一个变量趋于 $+\infty$; 判断独立必须检查联合分布能否分解为两个边缘分布的乘积. 使用前先确认端点开闭和矩形增量非负.

3.2 二维离散型随机变量

3.2.1 联合分布律、边缘与条件分布

定义 3.9 (二维离散型联合分布律)

若 (X, Y) 只能取有限对或可列多对值 (x_i, y_j) , 记

$$p_{ij} = P(X = x_i, Y = y_j),$$

则 $\{p_{ij}\}$ 称为联合分布律, 且

$$p_{ij} \geq 0, \quad \sum_i \sum_j p_{ij} = 1.$$

离散型问题的统一动作是: 列出满足目标条件的所有格点, 再把这些格点的概率相加.

定理 3.10 (由联合分布律得到边缘分布律)

若 $p_{ij} = P(X = x_i, Y = y_j)$, 则

$$p_{i\cdot} = P(X = x_i) = \sum_j p_{ij}, \quad p_{\cdot j} = P(Y = y_j) = \sum_i p_{ij}.$$

证明: 证明目标: 由联合概率推出每个分量自己的概率.

出发点: 固定 $X = x_i$ 后, Y 必取某个 y_j , 这些事件两两互斥.

关键转换: 把边缘事件按另一个变量的全部可能取值分割.

正式推导:

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\},$$

且右侧事件互斥, 所以

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}.$$

同理, $P(Y = y_j) = \sum_i p_{ij}$.

证明小结: “求谁的边缘, 就固定谁, 把另一个变量的所有情况加完”. 这是全概率思想在表格中的直接应用. ■

定义 3.11 (离散型条件分布律)

若 $p_{\cdot j} = P(Y = y_j) > 0$, 则

$$P(X = x_i | Y = y_j) = \frac{p_{ij}}{p_{\cdot j}}.$$

若 $p_{i\cdot} > 0$, 则

$$P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i\cdot}}.$$

分母为零时, 该条件分布没有定义.

证明: 证明目标: 由联合分布和边缘分布构造条件分布, 并验证它是合法分布.

出发点: 条件概率公式 $P(A | B) = P(AB)/P(B)$, 要求 $P(B) > 0$.

关键转换: 令 $A = \{X = x_i\}$, $B = \{Y = y_j\}$.

正式推导:

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}}.$$

各项非负, 且

$$\sum_i \frac{p_{ij}}{p_{\cdot j}} = \frac{\sum_i p_{ij}}{p_{\cdot j}} = 1.$$

证明小结: 条件分布就是把表格中固定的一行或一列重新归一化. 先查分母是否正, 是不可省略的定义条件. ■

3.2.2 离散型独立性

定理 3.12 (离散型独立判据)

二维离散型随机变量 X, Y 相互独立, 当且仅当对所有 i, j ,

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}.$$

只验证一两个格点不能证明独立; 找到一个反例格点即可否定独立.

证明: 证明目标: 把联合分布函数的独立性转化为表格中逐格相乘.

出发点: 独立事件满足联合概率等于边缘概率乘积.

关键转换: 对每个单点事件 $\{X = x_i\}$ 与 $\{Y = y_j\}$ 应用独立性.

正式推导: 若 X, Y 独立, 则

$$p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) = p_{i\cdot}p_{\cdot j}.$$

反之, 若逐格等式成立, 则任意集合 A, B 上

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x_i \in A} \sum_{y_j \in B} p_{ij} \\ &= \left(\sum_{x_i \in A} p_{i\cdot} \right) \left(\sum_{y_j \in B} p_{\cdot j} \right) \\ &= P(X \in A)P(Y \in B), \end{aligned}$$

故 X, Y 独立.

证明小结: 逐格乘积不仅是必要条件, 也是充分条件; 原因是所有离散事件都能由格点概率求和得到. ■

例题 3.13 (抽球: 联合、边缘、条件与独立性)

盒中有 2 个黑球、2 个红球、2 个白球, 不放回随机取 2 球. 令 X 为黑球数, Y 为红球数.

总取法为 $\binom{6}{2} = 15$. $X, Y \in \{0, 1, 2\}$, 且 $X + Y \leq 2$. 联合分布律为

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{1}{15}$
$X = 1$	$\frac{4}{15}$	$\frac{1}{15}$	0
$X = 2$	$\frac{1}{15}$	0	0

例如

$$P(X + Y \leq 1) = \frac{1}{15} + \frac{4}{15} + \frac{4}{15} = \frac{3}{5}.$$

X 的边缘分布律由逐行相加得到

$$P(X = 0) = \frac{2}{5}, \quad P(X = 1) = \frac{8}{15}, \quad P(X = 2) = \frac{1}{15}.$$

条件概率

$$P(Y = 1 | X = 0) = \frac{4/15}{6/15} = \frac{2}{3}.$$

又因

$$P(X = 2, Y = 2) = 0 \neq P(X = 2)P(Y = 2) = \frac{1}{225},$$

所以 X, Y 不独立.

结果检查: 九个格点概率之和为 1; 不放回造成取值约束 $X + Y \leq 2$, 这也解释了独立性为何失败.

注: [表格题的固定顺序] 先列 X, Y 的全部可能取值, 再逐格翻译事件并计算联合概率; 随后才能安全地求范围概率、边缘、条件和独立性. 不要先凭直觉判断独立, 也不要漏写概率为零的不可可能格点.

小结

二维离散型分布由“取值对加联合概率”组成. 求边缘就是按行或列求和; 求条件分布是固定一行或一列后重新归一化; 判断独立需逐格比较联合概率与两个边缘概率的乘积. 做题前先列全格点并检查总和为 1.

3.3 二维连续型随机变量

3.3.1 联合密度与区域概率

定义 3.14 (二维连续型随机变量与联合密度)

若存在非负可积函数 $f(x, y)$, 使联合分布函数可写成

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

则称 (X, Y) 为二维连续型随机变量, f 为联合概率密度.

定理 3.15 (联合密度的充要条件与区域概率)

函数 f 能作为某个二维连续型随机变量的联合密度, 当且仅当

$$f(x, y) \geq 0, \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx \, dy = 1.$$

对平面区域 G ,

$$P((X, Y) \in G) = \iint_G f(x, y) dx dy.$$

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

注: [二维连续型题的总口诀] “在哪个区域求概率, 就在哪个区域对联合密度积分.” 概率论的关键任务是画对**联合密度支持集**与**目标事件区域**的交集; 之后的积分次序可以按区域形状选择.

例题 3.16 (归一化、区域概率与联合分布函数)

设

$$f(x, y) = \begin{cases} Cxy, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{其他.} \end{cases}$$

求 C 、 $P(X + Y < 1)$ 及联合分布函数在 $0 < x < 1, 0 < y < 1$ 时的表达式.

先由归一性确定参数:

$$1 = C \int_0^1 x dx \int_0^1 y dy = \frac{C}{4}, \quad C = 4.$$

事件 $X + Y < 1$ 是单位正方形中的下三角形, 因此

$$P(X + Y < 1) = \int_0^1 \int_0^{1-x} 4xy dy dx = \frac{1}{6}.$$

当 $0 < x < 1, 0 < y < 1$ 时, 左下矩形完全位于支持集内, 故

$$F(x, y) = \int_0^x \int_0^y 4uv dv du = x^2 y^2.$$

完整分布函数还需按 $x \leq 0$ 、 $0 < x < 1$ 、 $x \geq 1$ 与 y 的三段组合讨论.

检查: 由 $f(x, y) = 4xy = (2x)(2y)$ 可预见 X, Y 独立, 且边缘分布函数均为区间内的平方函数.

3.3.2 边缘概率密度

定理 3.17 (由联合密度求边缘密度)

若 (X, Y) 的联合密度为 $f(x, y)$, 则

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

证明: 证明目标: 从联合密度推出单个分量的密度.

出发点: $F_X(x) = F(x, +\infty)$, 密度是分布函数的导数.

关键转换: 让 Y 遍历全部可能取值, 再对 x 求导.

正式推导:

$$F_X(x) = \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f(u, y) dy \right] du.$$

在可求导处,

$$f_X(x) = F'_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy.$$

f_Y 同理.

证明小结: 边缘化的本质与离散型完全相同: 把另一个变量的全部可能性“加掉”; 离散用求和, 连续用积分. ■

例题 3.18 (三角形支持集上的边缘密度)

设

$$f(x, y) = \begin{cases} kxy, & 0 \leq x \leq y \leq 1, \\ 0, & \text{其他.} \end{cases}$$

求 k 及两个边缘密度.

支持集是 $0 \leq x \leq y \leq 1$. 归一化得

$$1 = k \int_0^1 \int_x^1 xy dy dx = \frac{k}{8}, \quad k = 8.$$

固定 x 后, y 从 x 到 1:

$$f_X(x) = \int_x^1 8xy dy = 4x(1 - x^2), \quad 0 < x < 1.$$

固定 y 后, x 从 0 到 y :

$$f_Y(y) = \int_0^y 8xy \, dx = 4y^3, \quad 0 < y < 1.$$

区间外均为 0.

关键转折: 积分限不是背出来的, 而是固定一个坐标后, 从三角形中读出另一个坐标的范围.

3.3.3 条件概率密度与条件概率

定义 3.19 (条件概率密度)

在 $f_Y(y) > 0$ 时, 给定 $Y = y$ 条件下 X 的条件密度为

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

在 $f_X(x) > 0$ 时,

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

注: [为什么不能直接套普通条件概率] 连续型随机变量通常满足 $P(Y = y) = 0$, 所以不能把 $P(X \leq x, Y = y)/P(Y = y)$ 当作普通比值. 条件密度可理解为让 Y 落在 $[y, y + h]$ 后令 $h \rightarrow 0$ 得到的极限; 这是 B 级证明, 只需掌握“小区间条件化后归一化”的骨架.

证明: **证明目标:** 验证由联合密度与边缘密度构造出的条件密度是合法的一维密度.

出发点: 假设 $f_Y(y) > 0$, 且 $f(x, y) \geq 0$.

关键转换: 固定 y , 只对 x 积分.

正式推导:

$$f_{X|Y}(x | y) \geq 0,$$

并且

$$\int_{-\infty}^{+\infty} f_{X|Y}(x | y) \, dx = \frac{\int_{-\infty}^{+\infty} f(x, y) \, dx}{f_Y(y)} = 1.$$

证明小结: 条件密度是固定切片后除以该切片的总质量; 分母必须为正. 若分母为零, 公式没有定义. ■

例题 3.20 (“等号条件”与“区间条件”选择不同方法)

若题目问 $P(Y \leq b \mid X = x_0)$, 先求 $f_{Y|X}(y \mid x_0)$, 再在 $(-\infty, b]$ 上积分. 若题目问 $P(Y \leq b \mid X > a)$, 条件事件有正概率, 直接用

$$\frac{P(Y \leq b, X > a)}{P(X > a)}$$

更直接, 不必构造条件密度.

3.3.4 连续型独立性

定理 3.21 (连续型独立判据)

二维连续型随机变量 X, Y 相互独立, 当且仅当

$$f(x, y) = f_X(x)f_Y(y)$$

在平面上几乎处处成立.

证明: 证明目标: 用联合密度是否分解为边缘密度乘积判断独立.

出发点: 独立的定义是 $F(x, y) = F_X(x)F_Y(y)$.

关键转换: 对联合分布函数分别求偏导; 反向则积分恢复分布函数.

正式推导: 若 X, Y 独立, 在可求导处

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} [F_X(x)F_Y(y)] = f_X(x)f_Y(y).$$

反之, 若密度几乎处处分解, 则

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f_X(u)f_Y(v) \, dv \, du \\ &= F_X(x)F_Y(y), \end{aligned}$$

故 X, Y 独立.

证明小结: 除了代数因式分解, 还必须检查支持集是否能写成两个一维支持集的直积. 三角形或圆盘支持集通常已经否定独立. ■

注: [三种“看起来能拆”的误判]

1. 只看公式主体能拆, 忽略支持集不能拆;
2. 只在支持集内部验证乘积相等, 忽略外部;

3. 看到 $P(A \cap B)$ 就擅自写成 $P(A)P(B)$, 尚未证明独立.

小结

二维连续型概率由联合密度在平面区域上的二重积分得到. 求边缘是把另一个变量积掉; 求条件密度是联合除以相应边缘, 并先检查分母为正; 判断独立要同时检查密度乘积和支持集结构. 计算前先画支持集, 再决定积分限.

3.4 二维均匀分布与二维正态分布

3.4.1 二维均匀分布

定义 3.22 (二维均匀分布)

设 D 是面积为 S_D 的有界平面区域. 若

$$f(x, y) = \begin{cases} \frac{1}{S_D}, & (x, y) \in D, \\ 0, & \text{其他}, \end{cases}$$

则称 (X, Y) 在 D 上服从二维均匀分布.

对任意子区域 G ,

$$P((X, Y) \in G) = \frac{S_{G \cap D}}{S_D}.$$

面积比只适用于均匀分布; 一般联合密度仍须积分.

例题 3.23 (正方形与圆盘: 边缘和独立性的区别)

若 (X, Y) 在矩形 $[a, b] \times [c, d]$ 上均匀, 则 $X \sim U(a, b)$ 、 $Y \sim U(c, d)$, 且联合密度能分解, 故 X, Y 独立.

若 (X, Y) 在圆盘 $x^2 + y^2 \leq r^2$ 上均匀, 虽然图形对称, 但支持集不是两个区间的直积, X, Y 不独立; 其边缘密度也不是均匀密度. **对称不等于独立.**

3.4.2 二维正态分布

定义 3.24 (二维正态分布)

若联合密度为

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{x-\mu_1}{\sigma_1} \frac{y-\mu_2}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right]\right\},$$

其中 $\sigma_1, \sigma_2 > 0$, $|\rho| < 1$, 则记

$$(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho).$$

定理 3.25 (二维正态分布的常用性质)

若 (X, Y) 服从上述二维正态分布, 则:

1. $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$;
2. X, Y 独立当且仅当 $\rho = 0$;
3. 任意非零线性组合 $aX + bY$ 仍服从正态分布;
4. 条件分布仍为正态分布.

本阶段将这些结论作为 C 级结论使用, 不要求复现完整解析证明.

注: [边缘正态不保证联合正态] X, Y 各自服从正态分布, 并不能推出 (X, Y) 联合正态, 也不能推出 $X + Y$ 正态. 例如令 $Y = -X$, 二者边缘均可为标准正态, 但 $X + Y \equiv 0$. 若再给出 X, Y 独立, 则联合分布为二维正态, 线性组合才可直接套正态结论.

例题 3.26 (独立正态变量的装配问题)

木塞直径 $X \sim N(22.40, 0.03^2)$, 瓶口直径 $Y \sim N(22.50, 0.04^2)$, 且 X, Y 独立. 求木塞能装入瓶口的概率, 已知 $\Phi(2) = 0.9772$.

能装入的事件是 $X < Y$, 即 $X - Y < 0$. 独立正态的线性组合仍正态:

$$X - Y \sim N(22.40 - 22.50, 0.03^2 + 0.04^2) = N(-0.10, 0.05^2).$$

因此

$$P(X < Y) = P\left(\frac{X - Y + 0.10}{0.05} < 2\right) = \Phi(2) = 0.9772.$$

方法选择: 直接把二维事件化为一个正态线性组合, 比写二维积分简单得多.

小结

二维均匀分布把概率化为面积比, 但只能在密度恒定的支持集内使用; 矩形均匀分布的两个分量独立, 圆盘均匀分布通常不独立. 二维正态的边缘仍正态, 且只有在二维正态前提下, $\rho = 0$ 才与独立等价. 遇到独立正态的线性组合, 优先降成一维正态问题.

3.5 二维随机变量函数的分布

3.5.1 离散型: 合并产生同一函数值的概率

若 (X, Y) 为离散型, $Z = g(X, Y)$ 的分布律按两步得到: 先列出所有可能的 z , 再将所有满足 $g(x_i, y_j) = z$ 的格点概率相加:

$$P(Z = z) = \sum_{(i,j): g(x_i, y_j)=z} p_{ij}.$$

例题 3.27 (离散型和与最小值)

设联合分布律为

	$Y = -1$	$Y = 0$	$Y = 1$
$X = 0$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{5}{16}$
$X = 1$	$\frac{4}{16}$	$\frac{4}{16}$	0

求 $U = X + Y$ 与 $V = \min(X, Y)$ 的分布律.

U 的可能取值为 $-1, 0, 1, 2$. 逐条合并同一函数值:

u	-1	0	1	2
$P(U = u)$	$\frac{1}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	0

其中 $P(U = 0) = P(0, 0) + P(1, -1) = 6/16$, 其余同理.

V 的可能取值为 $-1, 0, 1$:

$$P(V = -1) = \frac{5}{16}, \quad P(V = 0) = \frac{11}{16}, \quad P(V = 1) = 0.$$

检查: 每张新分布表的概率之和都应为 1; 不能只代入格点而忘记合并重复函数值.

3.5.2 连续型通法: 分布函数法

定理 3.28 (二维随机变量函数的分布函数法)

设 (X, Y) 的联合密度为 $f(x, y)$, $Z = g(X, Y)$. 则

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

积分区域还必须与联合密度的支持集取交集. 若 F_Z 可导, 则 $f_Z(z) = F'_Z(z)$.

证明: 证明目标: 由 (X, Y) 的联合密度求 $Z = g(X, Y)$ 的分布.

出发点: 分布函数定义 $F_Z(z) = P(Z \leq z)$ 与区域概率公式.

关键转换: 把一维事件 $\{Z \leq z\}$ 翻译成平面事件 $\{g(X, Y) \leq z\}$.

正式推导:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(g(X, Y) \leq z) \\ &= \iint_{\{(x, y): g(x, y) \leq z\}} f(x, y) dx dy. \end{aligned}$$

若该分布函数在相应区间可导, 再求导得到密度.

证明小结: 分布函数法是通法. 最难的一步不是求导, 而是让曲线 $g(x, y) = z$ 在支持集上移动, 判断交集形状何时改变. ■

注: [分布函数法的固定流程]

1. 先由支持集估计 Z 的取值范围;
2. 写 $F_Z(z) = P(g(X, Y) \leq z)$;
3. 画支持集和边界曲线 $g(x, y) = z$;
4. 按交点变化处分段积分;
5. 求导后补写支持集外密度为 0, 并检查总积分为 1.

3.5.3 和的密度与卷积

定理 3.29 (和的密度公式)

设 (X, Y) 的联合密度为 $f(x, y)$, $Z = X + Y$, 则

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

若 X, Y 独立, 则

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx,$$

称为卷积公式.

证明: 证明目标: 从联合密度推导 $Z = X + Y$ 的密度.

出发点: 分布函数法与变上限积分求导.

关键转换: 事件 $X + Y \leq z$ 等价于 $y \leq z - x$.

正式推导:

$$F_Z(z) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-x} f(x, y) dy \right] dx.$$

对 z 求导得

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

独立时代入 $f(x, y) = f_X(x)f_Y(y)$, 即得卷积公式.

证明小结: 卷积的积分限表面上是全实轴, 真正有效的范围由 x 同时满足 $f_X(x) \neq 0$ 与 $f_Y(z-x) \neq 0$ 决定; 这就是支持集交集的几何意义. ■

例题 3.30 (两个独立均匀变量之和)

设 X, Y 相互独立且均服从 $U(0, 1)$, 求 $Z = X + Y$ 的密度.

由卷积公式,

$$f_Z(z) = \int_{-\infty}^{+\infty} \mathbf{1}_{(0,1)}(x)\mathbf{1}_{(0,1)}(z-x) dx.$$

有效的 x 同时满足 $0 < x < 1$ 与 $z - 1 < x < z$, 故积分区间长度为

$$f_Z(z) = \begin{cases} z, & 0 < z < 1, \\ 2 - z, & 1 \leq z < 2, \\ 0, & \text{其他.} \end{cases}$$

结果检查: 密度呈三角形、关于 $z = 1$ 对称, 总面积为 1; 这与 $Z \in (0, 2)$ 相符.

注: [其他变换公式的掌握层级] 对 $Z = aX + bY$ 、 $Z = Y/X$ 、 $Z = XY$ 可由变量变换得到带绝对值因子的积分公式. 一般多维变量变换的严格雅可比证明属于 C 级, 本阶段不要求复现; 遇到不熟悉的变换时, 优先使用分布函数法, 避免只背公式而漏掉支持集和绝对值.

3.5.4 正态线性组合

若 $X_i \sim N(\mu_i, \sigma_i^2)$ 相互独立, 则

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

正态线性组合应直接识别分布, 不必再做卷积或分布函数积分.

例题 3.31 (正态差的密度)

若 $X \sim N(1, \sigma^2)$ 、 $Y \sim N(0, 2\sigma^2)$ 独立, 令 $Z = X - Y$, 则

$$Z \sim N(1, 3\sigma^2),$$

所以

$$f_Z(z) = \frac{1}{\sqrt{6\pi}\sigma} \exp\left[-\frac{(z-1)^2}{6\sigma^2}\right].$$

小结

离散型函数分布靠“列新取值、合并同值格点”; 连续型通法是把 $Z \leq z$ 还原成平面区域后积分. 和的密度可用卷积, 但有效积分限仍由支持集决定; 正态独立线性组合直接写分布. 使用任何公式前先判断离散或连续、是否独立以及新变量的取值范围.

3.6 最大值、最小值与可加性

3.6.1 独立随机变量的最大值与最小值

定理 3.32 (最大值与最小值的分布)

设 $X_1, \dots, X_n$ 相互独立, 分布函数分别为 $F_1, \dots, F_n$. 令

$$M = \max(X_1, \dots, X_n), \quad L = \min(X_1, \dots, X_n).$$

则

$$F_M(z) = \prod_{i=1}^n F_i(z),$$

$$F_L(z) = 1 - \prod_{i=1}^n [1 - F_i(z)].$$

若还同分布于 F , 则

$$F_M(z) = [F(z)]^n, \quad F_L(z) = 1 - [1 - F(z)]^n.$$

证明: 证明目标: 从分布函数定义推出最大值、最小值公式.

出发点: 独立事件的交集概率等于概率乘积.

关键转换: “最大值不超过 z ” 等价于所有变量都不超过 z ; “最小值不超过 z ” 先取补集更容易.

正式推导:

$$\begin{aligned} F_M(z) &= P(M \leq z) \\ &= P(X_1 \leq z, \dots, X_n \leq z) \\ &= \prod_{i=1}^n P(X_i \leq z) = \prod_{i=1}^n F_i(z). \end{aligned}$$

对最小值,

$$\begin{aligned} F_L(z) &= 1 - P(L > z) \\ &= 1 - P(X_1 > z, \dots, X_n > z) \\ &= 1 - \prod_{i=1}^n [1 - F_i(z)]. \end{aligned}$$

证明小结: 最大值喜欢 “ $\leq$ ”, 最小值喜欢先转成 “ $>$ ”. 独立性是把交集概率拆成乘积的必要前提, 不能遗漏. ■

例题 3.33 (独立指数变量的最小值)

若 $X_i \sim E(\lambda_i)$ 相互独立, $L = \min(X_1, \dots, X_n)$, 则对 $z \geq 0$,

$$P(L > z) = \prod_{i=1}^n e^{-\lambda_i z} = e^{-(\lambda_1 + \dots + \lambda_n)z}.$$

所以

$$L \sim E(\lambda_1 + \dots + \lambda_n).$$

它描述多个独立部件中“第一个失效时间”.

3.6.2 常见可加性结论

在相互独立且参数条件匹配时, 有:

- $B(n_1, p) + B(n_2, p) \sim B(n_1 + n_2, p)$;
- $P(\lambda_1) + P(\lambda_2) \sim P(\lambda_1 + \lambda_2)$;
- 独立正态变量的线性组合仍为正态;
- 独立卡方变量之和的自由度相加;
- 独立指数变量的**最小值**仍为指数分布, 参数相加.

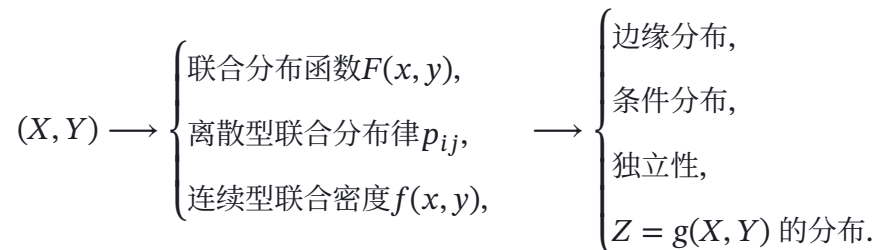
可加性必须连同独立性和“哪些参数必须相同”一起记, 不能只背分布名称.

小结

最大值分布直接研究“全部不超过”, 最小值分布先取补集研究“全部超过”; 只有相互独立时才能把联合事件拆成乘积. 常见分布的可加性还需核对参数条件, 例如二项分布必须有相同成功概率.

3.7 本章总结

3.7.1 知识结构



3.7.2 核心公式与适用条件

任务	公式	使用前检查
矩形概率	四项容斥	端点开闭
离散边缘	$p_{i\cdot} = \sum_j p_{ij}$	格点是否列全
连续边缘	$f_X(x) = \int f(x, y) dy$	支持集与积分限
条件密度	$f_{Y X} = f/f_X$	分母必须为正
独立判据	$p_{ij} = p_{i\cdot} p_{\cdot j}$ 或 $f = f_X f_Y$	对所有取值成立
函数分布	$F_Z(z) = P(g(X, Y) \leq z)$	新变量范围与分段点
卷积	$f_{X+Y}(z) = \int f_X(x) f_Y(z-x) dx$	独立性与有效积分区间
最大值	$F_M = \prod_i F_i$	相互独立
最小值	$F_L = 1 - \prod_i (1 - F_i)$	相互独立

3.7.3 常见题型与方法选择

- 联合分布函数题:** 把事件画成左下矩形, 有限矩形用四项容斥.
- 离散表格综合题:** 先列全格点, 再按“联合—边缘—条件—独立”顺序处理.
- 连续区域概率:** 支持集与目标区域取交集, 在交集上二重积分.
- 边缘与条件密度:** 先画切片范围; 条件密度用联合除边缘.
- 独立性:** 先看支持集能否拆成直积, 再验证联合是否等于边缘乘积.
- 函数分布:** 分布函数法是通法; 和可用卷积; 独立正态线性组合直接识别.
- 极值分布:** 最大值用全部“ $\leq$ ”, 最小值转为全部“ $>$ ”.

3.7.4 重点辨析

概念	本质区别	常见误判
联合与边缘	联合保留共同变化, 边缘忽略另一变量	边缘相同就以为联合相同
互斥与独立	互斥是不同时发生, 独立是不影响概率	把不可能格点当成独立证据
等号条件与区间条件	前者用条件密度, 后者可用普通条件概率	用 $0/0$ 直接定义条件概率
均匀与独立	区域内密度恒定不保证支持集可拆	圆盘均匀也判断独立
不相关与独立	一般不等价; 二维正态中才有特殊等价	看到 $\rho = 0$ 就无条件说独立
卷积与通法	卷积适合和且常需独立, 分布函数法通用	不画支持集直接写全实轴积分

3.7.5 易错点

1. 联合分布函数的矩形增量符号写错, 或漏检二重单调性.
2. 求边缘时积分限机械写成常数, 没有随支持集切片变化.
3. 条件密度的分母选错, 或没有检查边缘密度是否大于零.
4. 仅凭联合密度主体可因式分解就判断独立, 忽略支持集.
5. 一般二维均匀分布擅用面积比以外的“独立”结论.
6. 只知道两个边缘正态, 就断言线性组合正态.
7. 求 $Z = g(X, Y)$ 时不先求取值范围, 导致漏段或积分限错误.
8. 最大值、最小值公式忘记相互独立前提.

3.7.6 证明掌握清单

- **A 级:** 会独立推导边缘分布、条件分布与独立判据; 会用分布函数法和卷积求函数分布; 会推导最大值、最小值公式.
- **B 级:** 理解条件密度来自小区间条件化; 理解卷积积分限的几何意义; 理解二维正态中独立与相关的特殊关系.
- **C 级:** 一般多维变量变换的严格雅可比证明, 以及二维正态性质的完整解析证明, 本阶段只要求准确使用结论.

3.7.7 自测题

练习 3.34 (离散型综合)

给出一个 2×3 联合分布表, 其中含未知参数 a . 请依次求 a 、两个边缘分布、 $P(X+Y \leq 1)$ 、一个有定义的条件分布, 并判断独立性.

练习 3.35 (三角形支持集)

设

$$f(x, y) = \begin{cases} c, & 0 < x < y < 1, \\ 0, & \text{其他.} \end{cases}$$

求 c 、 f_X 、 f_Y 、 $f_{Y|X}$, 并判断 X, Y 是否独立.

练习 3.36 (函数分布)

设 X, Y 独立且均服从 $U(0, 1)$. 不用背结论, 分别用分布函数法求 $\max(X, Y)$ 与 $\min(X, Y)$ 的分布函数, 并解释为何两式结构不同.

练习 3.37 (卷积与检查)

设 X, Y 独立且都服从参数为 1 的指数分布. 用卷积求 $Z = X + Y$ 的密度, 再检查其非负性、支持集和归一性.

注: [本章自检] 完成本章后, 应能回答: 联合、边缘和条件分布分别保留什么信息? 为什么连续型条件 $X = x$ 不能直接套普通条件概率? 如何从支持集快速排除独立? 为什么分布函数法是函数分布的通法? 最大值与最小值公式中的独立性具体用在了哪一步?

第四章 随机变量的数字特征

本章导读

前几章用分布完整描述随机变量, 但很多问题不需要保存全部分布信息. 例如, 比较两种方案时可能只关心长期平均收益、波动大小以及两个指标是否同向变化. 数字特征就是把分布压缩成少数有解释力的数:

中心位置: 期望, 波动程度: 方差, 共同变化: 协方差与相关系数.

本章始终沿着同一条计算链工作:

已知分布 $\longrightarrow E(X), E(X^2), E(XY) \longrightarrow D(X), \text{Cov}(X, Y), \rho_{XY}$.

其中期望是根. 方差、协方差和相关系数最终都要回到期望计算.

4.1 数学期望

4.1.1 定义、存在性与直观含义

定义 4.1 (离散型随机变量的数学期望)

设离散型随机变量 X 的分布律为

$$P(X = x_k) = p_k, \quad k = 1, 2, \dots$$

若级数 $\sum_k x_k p_k$ 绝对收敛, 则定义

$$E(X) = \sum_k x_k p_k.$$

否则称 X 的数学期望不存在.

定义 4.2 (连续型随机变量的数学期望)

设连续型随机变量 X 的密度为 $f(x)$. 若

$$\int_{-\infty}^{+\infty} |x|f(x) dx < +\infty,$$

则定义

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

数学期望是按概率加权的平均位置. 它不一定是随机变量可能取到的值, 也不保证每个随机变量都存在期望. 绝对收敛保证求和或积分次序不影响唯一结果.

注: [对称不能代替存在性检查] 标准柯西密度

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

关于原点对称, 但

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

因此其期望不存在. 把左右两端对称抵消得到的主值为 0, 不是概率论定义下的期望.

例题 4.3 (离散型期望与函数期望)

设

x	-2	0	2
$P(X=x)$	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{3}{10}$

求 $E(X)$ 与 $E(3X^2 + 5)$.

直接按“取值乘概率”求和:

$$E(X) = (-2)\frac{4}{10} + 0 \cdot \frac{3}{10} + 2\frac{3}{10} = -\frac{1}{5}.$$

先求

$$E(X^2) = 4\frac{4}{10} + 0 + 4\frac{3}{10} = \frac{14}{5},$$

再利用期望的线性性:

$$E(3X^2 + 5) = 3E(X^2) + 5 = \frac{67}{5}.$$

方法选择: 无需先求 $3X^2 + 5$ 的分布律; 直接计算函数期望更短.

4.1.2 随机变量函数的期望

定理 4.4 (一维随机变量函数的期望)

若 $Y = g(X)$ 且相应绝对收敛条件成立, 则:

- 离散型:

$$E[g(X)] = \sum_k g(x_k) p_k;$$

- 连续型:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

这条公式常称为“无须先求函数分布的期望公式”. 口诀是: 求谁的期望, 就把谁乘在原分布前面求和或积分.

例题 4.5 (分段函数的期望)

设 $X \sim E(1)$, 令 $Z = \max(X, 1)$, 求 $E(Z)$.

Z 在 $0 < X < 1$ 时等于 1, 在 $X \geq 1$ 时等于 X , 故

$$\begin{aligned} E(Z) &= \int_0^1 1 \cdot e^{-x} dx + \int_1^{+\infty} x e^{-x} dx \\ &= (1 - e^{-1}) + 2e^{-1} = 1 + e^{-1}. \end{aligned}$$

关键转折: 先判断 Z 在不同区间的表达式, 再把正确的函数乘密度; 不能把 $\max(X, 1)$ 误当成连续型随机变量处处都有普通变换密度后才计算.

4.1.3 二维随机变量函数的期望

定理 4.6 (二维函数期望公式)

设 $Z = g(X, Y)$, 相应绝对收敛条件成立.

- 若 (X, Y) 离散, 联合分布律为 p_{ij} , 则

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) p_{ij}.$$

- 若 (X, Y) 连续, 联合密度为 $f(x, y)$, 则

$$E[g(X, Y)] = \iint_{\mathbb{R}^2} g(x, y)f(x, y) \, dx \, dy.$$

例题 4.7 (二维离散型函数期望)

设联合分布律为

	$X = 1$	$X = 2$	$X = 3$
$Y = -1$	0.2	0.1	0
$Y = 0$	0.1	0	0.3
$Y = 1$	0.1	0.1	0.1

若 $Z = (Y - X)^2$, 则

$$E(Z) = \sum_{x,y} (y - x)^2 P(X = x, Y = y).$$

逐格代入即可, 无需先把相同的 Z 值合并成新分布律.

例题 4.8 (相约问题的平均等待时间)

甲乙约在 8–12 点见面, 二人到达时刻 X, Y 独立且均服从 $U(0, 4)$. 先到者等待时间为 $|X - Y|$, 求平均等待时间.

联合密度在正方形 $(0, 4)^2$ 上为 $1/16$. 由对称性只积上三角形再乘 2:

$$\begin{aligned} E|X - Y| &= 2 \int_0^4 \int_0^y (y - x) \frac{1}{16} \, dx \, dy \\ &= \frac{4}{3}. \end{aligned}$$

所以平均等待时间为 $4/3$ 小时, 即 80 分钟.

结果检查: 等待时间在 $[0, 4]$ 内, 平均值 $4/3$ 合理; 对称性只减少计算量, 不改变联合密度.

4.1.4 期望的运算性质

定理 4.9 (期望的常用性质)

在相关期望存在时:

1. $E(C) = C$;
2. $E(aX + bY + c) = aE(X) + bE(Y) + c$;
3. 更一般地, $E(\sum_i a_i X_i + c) = \sum_i a_i E(X_i) + c$;
4. 若 X, Y 独立, 则 $E(XY) = E(X)E(Y)$.

注: [线性无条件, 乘积拆分有条件] 期望的加减可以无条件拆开, 不要求独立; 乘积期望拆成期望乘积则需要独立, 或至少需要后文所说的“不相关”. 这是本章最常见的条件混淆.

注: [证明等级] 期望线性性质在离散、连续情形下逐条验证属于 C 级, 本阶段直接熟练使用; 但遇到具体题目要求证明某个期望等于指定值时, 必须从分布、函数期望公式或线性性质完整写出计算链.

小结

期望是按概率加权的平均位置. 使用前先检查绝对收敛; 求函数期望通常直接把函数乘原分布, 无需先求新分布. 二维情形把求和推广为双重求和、把积分推广为二重积分. 期望线性无须独立, 只有乘积期望的拆分需要额外条件.

4.2 方差与标准差

4.2.1 定义与计算公式

定义 4.10 (方差与标准差)

若 $E[(X - E(X))^2]$ 存在, 则定义

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

为 X 的方差, 定义

$$\sigma_X = \sqrt{D(X)}$$

为标准差.

方差衡量随机变量围绕期望的波动程度. 方差单位是原单位的平方; 标准差与 X 同量纲, 更便于直观解释.

定理 4.11 (方差的易计算公式)

若 $E(X^2)$ 存在, 则

$$D(X) = E(X^2) - [E(X)]^2.$$

注: [公式为什么成立 (B 级)] 把 $E(X)$ 视为常数, 展开

$$E[(X - E(X))^2] = E(X^2) - 2E(X)E(X) + [E(X)]^2,$$

即可得到结论. 真正起作用的是期望的线性性.

证明: **证明目标:** 根据定义验证给定随机变量的方差.

出发点: $D(X) = E(X^2) - [E(X)]^2$.

关键转换: 先从分布分别计算一阶矩和二阶矩.

正式推导: 若离散型 X 的分布律为 p_k , 则

$$E(X) = \sum_k x_k p_k, \quad E(X^2) = \sum_k x_k^2 p_k,$$

所以

$$D(X) = \sum_k x_k^2 p_k - \left(\sum_k x_k p_k \right)^2.$$

连续型时把求和替换为对密度的积分即可.

证明小结: 验证方差的固定路线是“一阶矩、二阶矩、相减”, 最后检查结果非负. ■

例题 4.12 (由函数分布求方差)

设 $X \sim U(-3, 2)$, 令

$$Y = \begin{cases} 1, & X \geq 0, \\ -1, & X < 0. \end{cases}$$

求 $D(Y)$.

区间长度为 5, 故

$$P(Y = 1) = P(X \geq 0) = \frac{2}{5}, \quad P(Y = -1) = \frac{3}{5}.$$

于是

$$E(Y) = \frac{2}{5} - \frac{3}{5} = -\frac{1}{5}, \quad E(Y^2) = 1.$$

所以

$$D(Y) = 1 - \left(-\frac{1}{5}\right)^2 = \frac{24}{25}.$$

检查: Y 只取 ± 1 , 方差不超过 1, 结果合理.

4.2.2 方差的运算性质

定理 4.13 (方差的常用性质)

1. $D(C) = 0$;
2. $D(X + C) = D(X)$;
3. $D(aX) = a^2 D(X)$;
4. 若 X, Y 独立, 则

$$D(X \pm Y) = D(X) + D(Y).$$

更一般的非独立情形将在协方差一节补全.

注: [和差方差的三个易错点] 常数平移在方差中消失; 系数提出要平方; 独立时无论是 $X + Y$ 还是 $X - Y$, 两个方差之间都是加号. 若未给独立或不相关, 不能擅自拆开.

定理 4.14 (方差为零的含义)

$D(X) = 0$ 当且仅当存在常数 c , 使 $P(X = c) = 1$. 该严格证明属于 C 级, 本阶段准确记住结论并会在题目中调用.

小结

方差是到期望的平方偏差的平均, 标准差与原变量同量纲. 计算时优先用“二阶矩减一阶矩平方”. 平移不改变方差, 缩放系数要平方; 和差方差只有在独立或不相关时才能直接相加.

4.3 常见分布的期望与方差

4.3.1 常用结论表

分布	参数与记号	$E(X)$	$D(X)$
0-1 分布	$P(X = 1) = p$	p	$p(1 - p)$
二项分布	$B(n, p)$	np	$np(1 - p)$
泊松分布	$P(\lambda)$	λ	λ
几何分布	$Ge(p)$, 取值从 1 起	$1/p$	$(1 - p)/p^2$
均匀分布	$U(a, b)$	$(a + b)/2$	$(b - a)^2/12$
指数分布	$E(\lambda)$	$1/\lambda$	$1/\lambda^2$
正态分布	$N(\mu, \sigma^2)$	μ	σ^2
卡方分布	$\chi^2(n)$	n	$2n$

必须连同参数口径一起记. 例如泊松分布的期望和方差等于 λ ; 指数分布的期望与方差却分别是 $1/\lambda$ 与 $1/\lambda^2$.

例题 4.15 (由期望和方差反求二项参数)

已知 $X \sim B(n, p)$, 且

$$E(X) = 2.4, \quad D(X) = 1.44.$$

求 n, p .

由

$$np = 2.4, \quad np(1 - p) = 1.44,$$

两式相除得

$$1 - p = \frac{1.44}{2.4} = 0.6, \quad p = 0.4.$$

进而 $n = 2.4/0.4 = 6$.

检查: n 是正整数、 $0 < p < 1$, 且代回方差得到 $6 \times 0.4 \times 0.6 = 1.44$.

例题 4.16 (由方差反求二阶矩)

若 $X \sim P(\lambda)$ 且 $E(X^2) = 6$, 则

$$E(X^2) = D(X) + [E(X)]^2 = \lambda + \lambda^2 = 6.$$

由 $\lambda > 0$ 得 $\lambda = 2$. 当期望、方差有现成结论时, 不必重新对分布律求和.

4.3.2 独立正态的线性组合

若 $X_i \sim N(\mu_i, \sigma_i^2)$ 相互独立, 则

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

例题 4.17 (现场确定正态参数)

设 $X \sim N(-3, 1)$ 、 $Y \sim N(2, 1)$, 且相互独立. 令 $Z = X - 2Y + 7$, 则

$$E(Z) = -3 - 2 \times 2 + 7 = 0,$$

$$D(Z) = 1 + (-2)^2 \times 1 = 5.$$

因此 $Z \sim N(0, 5)$.

4.3.3 最大值与最小值的期望技巧

对任意实数 x, y , 有

$$\max(x, y) + \min(x, y) = x + y, \quad \max(x, y) \min(x, y) = xy,$$

以及

$$\max(x, y) = \frac{x + y + |x - y|}{2}, \quad \min(x, y) = \frac{x + y - |x - y|}{2}.$$

这些恒等式常比先求极值分布再积分更短.

例题 4.18 (两个独立指数变量的极值期望)

设 X, Y 独立且均服从 $E(1)$. 令

$$U = \max(X, Y), \quad V = \min(X, Y).$$

求 $E(U), E(V), E(UV)$.

独立指数变量的最小值仍为指数分布, 参数相加, 因此

$$V \sim E(2), \quad E(V) = \frac{1}{2}.$$

又由 $U + V = X + Y$,

$$E(U) = E(X) + E(Y) - E(V) = 1 + 1 - \frac{1}{2} = \frac{3}{2}.$$

由 $UV = XY$ 且 X, Y 独立,

$$E(UV) = E(XY) = E(X)E(Y) = 1.$$

易错点: U, V 一般不独立, 不能把 $E(UV)$ 直接拆成 $E(U)E(V)$; 这里是先用 $UV = XY$ 换回独立的 X, Y .

小结

常见分布的期望、方差应连同参数口径熟记. 反求参数时联立期望和方差; 已知期望、方差时可反求二阶矩. 独立正态线性组合直接由期望和方差确定参数; 极值期望优先检查和、积及绝对值恒等式, 避免不必要地先求完整分布.

4.4 切比雪夫不等式

定理 4.19 (切比雪夫不等式)

若 $E(X) = \mu$ 、 $D(X) = \sigma^2$ 存在, 则对任意 $\varepsilon > 0$,

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

等价地,

$$P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$

注: [为什么成立 (B 级证明骨架)] 在事件 $|X - \mu| \geq \varepsilon$ 上, $(X - \mu)^2 \geq \varepsilon^2$. 因此该事件的概率质量乘以 ε^2 不超过全部平方偏差的期望 $D(X)$. 离散型用求和、连续型用积分即可写成严格放缩. 本阶段应能复述这条思路并准确调用结论.

例题 4.20 (直接估计偏离期望的概率)

若 $D(X) = 2$, 则

$$P(|X - E(X)| \geq 2) \leq \frac{2}{2^2} = \frac{1}{2}.$$

切比雪夫不等式只利用期望和方差, 不需要知道 X 的具体分布, 因此给出的是普适上界, 不一定精确.

例题 4.21 (先构造中心化变量再使用不等式)

已知

$$E(X) = E(Y) = 2, \quad D(X) = 1, \quad D(Y) = 4, \quad \rho_{XY} = 0.5.$$

估计 $P(|X - Y| \geq 6)$ 的上界.

令 $Z = X - Y$. 先检查中心:

$$E(Z) = E(X) - E(Y) = 0,$$

所以 $|X - Y| = |Z - E(Z)|$. 由

$$\text{Cov}(X, Y) = \rho_{XY} \sqrt{D(X)D(Y)} = 0.5 \times 1 \times 2 = 1,$$

得

$$D(Z) = D(X) + D(Y) - 2 \text{Cov}(X, Y) = 1 + 4 - 2 = 3.$$

因此

$$P(|X - Y| \geq 6) \leq \frac{3}{36} = \frac{1}{12}.$$

方法选择: 切比雪夫形式必须围绕期望; 若题面没有显式出现期望, 应先构造新变量并计算其期望、方差.

小结

切比雪夫不等式用方差控制偏离期望的概率, 不依赖具体分布. 使用时先确认期望、方差存在, 并把目标事件写成 $|Z - E(Z)| \geq \varepsilon$; 右端一定是“方差除以距离平方”.

4.5 协方差

4.5.1 定义、计算与性质

定义 4.22 (协方差)

若相应期望存在, 定义

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

为 X, Y 的协方差.

协方差为正, 表示两个变量相对各自期望的偏差倾向同号; 协方差为负, 表示偏差倾向异号. 但它受单位和尺度影响, 不能直接跨量纲比较强弱.

定理 4.23 (协方差的易计算公式)

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

注: [展开思路 (B 级)] 把定义中的乘积展开并使用期望线性性, 四项中有三项合并, 得到“乘积期望减期望乘积”. 因此从联合分布计算协方差时, 固定求 $E(XY), E(X), E(Y)$ 三个量.

定理 4.24 (协方差的运算性质)

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
2. $\text{Cov}(X, C) = 0$;
3. $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$;
4. $\text{Cov}(X, X) = D(X)$;
5. 协方差对每个变量都具有线性性;
6. 若 X, Y 独立, 则 $\text{Cov}(X, Y) = 0$;
- 7.

$$D(X \pm Y) = D(X) + D(Y) \pm 2 \text{Cov}(X, Y).$$

证明: 证明目标: 推导一般和差方差公式, 并指出独立情形为何可简化.

出发点: 方差与协方差的定义.

关键转换: 把中心化后的和差平方展开.

正式推导:

$$\begin{aligned} D(X \pm Y) &= E\{[(X - E(X)) \pm (Y - E(Y))]^2\} \\ &= E[(X - E(X))^2] + E[(Y - E(Y))^2] \\ &\quad \pm 2E[(X - E(X))(Y - E(Y))] \\ &= D(X) + D(Y) \pm 2 \text{Cov}(X, Y). \end{aligned}$$

若 X, Y 独立, 则 $E(XY) = E(X)E(Y)$, 故协方差为零, 和差方差均化为方差之和.

证明小结: 方差不能无条件拆开, 缺少的“小尾巴”正是二倍协方差. 独立只是一种让协方差归零的充分条件. ■

例题 4.25 (由联合分布计算协方差)

设

	$Y = -1$	$Y = 0$	$Y = 1$
$X = 0$	0.07	0.18	0.15
$X = 1$	0.08	0.32	0.20

求 $\text{Cov}(X^2, Y^2)$.

由于 $X \in \{0, 1\}$, 有 $X^2 = X$; 由于 $Y \in \{-1, 0, 1\}$, Y^2 是 $Y \neq 0$ 的指示变量.

$$E(X^2) = P(X = 1) = 0.60,$$

$$E(Y^2) = P(|Y| = 1) = 0.07 + 0.15 + 0.08 + 0.20 = 0.50,$$

$$E(X^2 Y^2) = P(X = 1, |Y| = 1) = 0.08 + 0.20 = 0.28.$$

所以

$$\text{Cov}(X^2, Y^2) = 0.28 - 0.60 \times 0.50 = -0.02.$$

关键转折: 先利用取值简化平方, 再把乘积期望识别成一个事件概率.

小结

协方差衡量两个变量中心化偏差的共同变化, 计算时固定求三个期望. 协方差对加法线性、常数消失, 且补全了非独立情形的和差方差公式. 独立推出协方差为零, 但逆命题一般不成立.

4.6 相关系数

协方差能够判断共同变化的方向, 但它会随单位和尺度改变. 例如把“米”换成“厘米”, 协方差会被放大. 为了得到一个无量纲、便于比较的指标, 需要把协方差标准化.

定义 4.26 (相关系数)

设随机变量 X, Y 的方差存在且

$$D(X) > 0, \quad D(Y) > 0.$$

称

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为 X, Y 的相关系数.

注: 定义前必须检查两个方差都严格大于零. 若某个变量几乎处处为常数, 它没有随机波动, 分母为零, 相关系数便没有定义. 相关系数描述的是**线性相关程度**, 不是一般意义下所有依赖关系的强弱.

定理 4.27 (相关系数的基本性质)

若 $D(X) > 0, D(Y) > 0$, 则:

1. $\rho_{XY} = \rho_{YX}$;
2. $-1 \leq \rho_{XY} \leq 1$;
3. 对 $a, c \neq 0$,

$$\rho_{aX+b, cY+d} = \text{sgn}(ac)\rho_{XY};$$

4. $|\rho_{XY}| = 1$ 当且仅当存在常数 $a \neq 0, b$, 使

$$Y = aX + b \quad \text{几乎处处成立.}$$

此时 $\rho_{XY} = 1$ 对应 $a > 0$, $\rho_{XY} = -1$ 对应 $a < 0$.

注: [为什么相关系数的绝对值不超过 1] 令

$$U = \frac{X - E(X)}{\sqrt{D(X)}}, \quad V = \frac{Y - E(Y)}{\sqrt{D(Y)}}.$$

则 $E(U) = E(V) = 0, E(U^2) = E(V^2) = 1$, 并且 $\rho_{XY} = E(UV)$. 由柯西-施瓦茨不等式,

$$|E(UV)|^2 \leq E(U^2)E(V^2) = 1.$$

所以 $|\rho_{XY}| \leq 1$. 等号成立恰好对应 U, V 几乎处处成比例, 从而 X, Y 存在几乎处处成立的线性关系.

这里使用了较高层次的不等式及其等号条件. 现阶段应掌握标准化思路和证明骨架; 严格的不等式理论细节本阶段不要求复现.

例题 4.28 (完全负线性相关)

在长度为 1 的木棍上随机选一点截断, 令左段长度为 X , 右段长度为 Y . 求 ρ_{XY} .

由几何关系, $Y = 1 - X$. 只要截点不是固定的, 就有 $D(X) > 0$, 且

$$\text{Cov}(X, Y) = \text{Cov}(X, 1 - X) = -D(X),$$

$$D(Y) = D(1 - X) = D(X).$$

因此

$$\rho_{XY} = \frac{-D(X)}{\sqrt{D(X)}\sqrt{D(X)}} = -1.$$

方法选择: 看到确定的线性关系, 直接使用 $|\rho| = 1$ 的性质, 比先求联合分布更简洁.

结果检查: 左段变长时右段必等量变短, 完全反向线性变化, 故结果应为 -1 .

例题 4.29 (尺度变化不改变相关强弱)

若 $D(X), D(Y) > 0$ 且 $\rho_{XY} = 0.8$, 令 $U = 100X + 3$, $V = -2Y + 5$, 则

$$\rho_{UV} = \text{sgn}(100 \times (-2))\rho_{XY} = -0.8.$$

平移不影响相关系数; 正比例缩放不改变符号, 负比例缩放会反转符号.

协方差与相关系数的辨析

比较项	协方差	相关系数
定义	$\text{Cov}(X, Y)$	协方差除以两个标准差
量纲	有量纲, 随单位改变	无量纲
取值范围	没有统一范围	$[-1, 1]$
主要作用	判断共同变化方向, 补全和差方差	衡量线性相关的方向与标准化程度
使用条件	相关二阶矩存在	还要求两个方差严格大于零
常见误判	数值大就认为关系强	$\rho = 0$ 就认为完全没有关系

小结

相关系数是标准化后的协方差, 取值在 $[-1, 1]$, 只衡量线性相关. 使用前先检查两个方差是否严格大于零; $|\rho| = 1$ 表示几乎处处存在确定的线性关系, 而 $\rho = 0$ 只表示没有线性相关, 不能直接推出独立.

4.7 不相关与独立

这一节解决本章最重要的逻辑辨析: 两个变量的协方差为零, 究竟能说明什么? 它与独立到底相差多远?

定义 4.30 (不相关)

若 X, Y 的二阶矩存在, 且

$$\text{Cov}(X, Y) = 0,$$

则称 X, Y 不相关. 当 $D(X) > 0, D(Y) > 0$ 时, 这也等价于 $\rho_{XY} = 0$.

定理 4.31 (不相关的等价刻画)

设 X, Y 的二阶矩存在, 则下列命题等价:

1. $\text{Cov}(X, Y) = 0$;
2. $E(XY) = E(X)E(Y)$;
3. $D(X + Y) = D(X) + D(Y)$;
4. $D(X - Y) = D(X) + D(Y)$.

若再有 $D(X) > 0, D(Y) > 0$, 它们还等价于 $\rho_{XY} = 0$.

证明: 证明目标: 证明“不相关”的几种常见说法可以互相替换.

出发点: 协方差易计算公式、和差方差公式以及相关系数定义.

关键转换: 每个等式与标准形式比较, 分离出其中的协方差项.

正式推导: 由

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y),$$

立即得到命题 1 与命题 2 等价. 又因为

$$D(X + Y) - D(X) - D(Y) = 2 \operatorname{Cov}(X, Y),$$

$$D(X - Y) - D(X) - D(Y) = -2 \operatorname{Cov}(X, Y),$$

所以命题 1 分别与命题 3、命题 4 等价. 若两个方差严格大于零, 则相关系数的分母为正, 故

$$\rho_{XY} = 0 \iff \operatorname{Cov}(X, Y) = 0.$$

证明小结: 这些等价关系都来自同一个量 $\operatorname{Cov}(X, Y)$. 使用时可选择题目中最容易计算的一种形式, 但必须先确认相关矩存在; 涉及相关系数时还要检查方差不为零. ■

定理 4.32 (独立与不相关的关系)

若 X, Y 相互独立且二阶矩存在, 则 X, Y 不相关. 反过来一般不成立.

证明: **证明目标:** 由独立性推出协方差为零.

出发点: 独立随机变量的函数乘积期望可拆分.

关键转换: 把 $E(XY)$ 拆为 $E(X)E(Y)$.

正式推导: 由于 X, Y 独立,

$$E(XY) = E(X)E(Y).$$

因此

$$\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0.$$

证明小结: 独立控制的是联合分布的整体结构, 所以能推出不相关; 不相关只控制一个混合二阶矩, 信息远少于独立, 因而通常不能反推. ■

例题 4.33 (不相关但不独立的反例)

设 $X \sim U(-1, 1)$, 令 $Y = X^2$. 判断 X, Y 是否相关、是否独立.

由于 X 关于原点对称,

$$E(X) = 0, \quad E(XY) = E(X^3) = 0.$$

于是

$$\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0,$$

所以 X, Y 不相关.

但 $Y = X^2$ 是由 X 完全决定的. 例如一旦知道 X , 便确定知道 Y , 二者显然不是独立的.
关键转折: 奇函数在对称区间上的期望为零, 只能说明线性共同变化被抵消, 不能消除确定的非线性关系.

易错提醒: “相关系数为零”应读作“无线性相关”, 不能读作“没有任何关系”.

定理 4.34 (二维正态分布的特殊结论)

若 (X, Y) 服从二维正态分布, 则

$$X, Y \text{ 相互独立} \iff \text{Cov}(X, Y) = 0 \iff \rho_{XY} = 0,$$

其中相关系数形式要求两个方差严格大于零.

注: 这是二维正态分布的特殊结构, 不能推广到任意联合分布. 做题时只有明确给出“联合正态”或“二维正态”, 才能由不相关反推独立; 仅知道两个边缘分布各自为正态仍不够.

独立与不相关的辨析

比较项	独立	不相关
本质	联合分布可由边缘分布分解	协方差为零
控制范围	所有适当函数的联合关系	只控制一个二阶线性指标
常用判定	联合分布函数或密度分解、事件乘法	$E(XY) = E(X)E(Y)$
一般关系	有矩条件下推出不相关	一般不能推出独立
特殊情形	二维正态中与不相关等价	需先确认联合正态

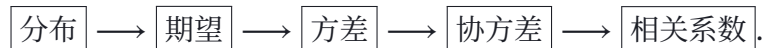
小结

不相关等价于协方差为零, 也等价于乘积期望可拆分以及和差方差可直接相加. 独立在矩存在时一定推出不相关, 但不相关通常不能推出独立; 只有在二维正态等特殊结构中才能反推. 判断时先问“题目给的是一般联合分布, 还是联合正态”.

4.8 本章总结

知识结构

本章可以按一条主线组织:



期望描述中心; 方差描述单个变量围绕中心的波动; 协方差描述两个变量共同变化; 相关系数把协方差标准化. 独立性属于联合分布层面的性质, 不相关只是二阶矩层面的性质.

核心公式与适用条件

对象	核心公式	使用前检查
离散型期望	$E(X) = \sum_k x_k p_k$	绝对收敛
连续型期望	$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$	绝对可积
函数期望	$E[g(X)] = \sum g(x_k) p_k$ 或 $\int g(x) f(x) dx$	原变量分布已知, 期望存在
方差	$D(X) = E(X^2) - [E(X)]^2$	二阶矩存在
和差方差	$D(X \pm Y) = D(X) + D(Y) \pm 2 \text{Cov}(X, Y)$	二阶矩存在; 不能漏协方差
协方差	$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$	相关矩存在
相关系数	$\rho = \text{Cov}(X, Y) / \sqrt{D(X)D(Y)}$	两个方差严格大于零
切比雪夫	$P(X - E(X) \geq \varepsilon) \leq D(X) / \varepsilon^2$	$\varepsilon > 0$, 期望和方差存在

常见题型与方法选择

1. **求函数的期望:** 已知原变量分布时, 优先直接计算 $E[g(X)]$, 不要先求新变量分布.
2. **求方差:** 通常先求 $E(X)$ 、 $E(X^2)$, 再用快捷公式; 系数提出时必须平方.
3. **参数反求:** 把题目给出的期望、方差代入对应分布公式, 再检查参数范围.
4. **求协方差:** 固定计算 $E(XY)$, $E(X)$, $E(Y)$; 若存在简单线性关系, 优先用运算性质.
5. **求线性组合方差:** 先写完整协方差公式, 只有确认独立或不相关后才能删去交叉项.
6. **概率粗界:** 只知道期望和方差时, 把事件整理为偏离期望的形式, 再使用切比雪夫不等式.
7. **判断独立:** 一般回到联合分布; 只有联合正态等明确特殊条件下, 才可用零相关反推独立.

重点辨析

容易混淆的说法	正确结论	常见误判
期望线性性	$E(aX + bY) = aE(X) + bE(Y)$, 不要求独立	看到两个变量就额外要求独立
乘积期望	独立可推出 $E(XY) = E(X)E(Y)$	无条件拆分乘积
方差线性性	缩放系数平方, 和差还含协方差	把方差当作期望直接拆
协方差与相关系数	前者有量纲, 后者标准化且有界	用协方差绝对值直接比较强弱
零相关与独立	独立推出零相关, 逆命题一般不成立	看见 $\rho = 0$ 就判独立
二维正态	联合正态下零相关与独立等价	两个边缘正态就套用该结论

易错点清单

- 不检查期望的绝对收敛或绝对可积, 误把发散的正负部分抵消成有限值;
- 求 $E[g(X)]$ 时误把密度写成 $g(f(x))$, 忘记“函数乘原密度”;
- 把 $D(aX + b)$ 写成 $aD(X) + b$;
- 未知独立性时擅自写 $D(X + Y) = D(X) + D(Y)$;
- 切比雪夫不等式没有围绕 $E(X)$ 整理事件, 或忘记分母平方;
- 相关系数分母为零时仍继续计算;
- 把“无相关”理解成“无任何依赖关系”;
- 忘记二维正态结论要求的是联合正态, 而非仅边缘正态.

证明掌握清单

- **A 级:** 能由定义计算并验证期望、方差、协方差和相关系数; 能完整推出和差方差公式, 并逐项说明条件.
- **B 级:** 能复述切比雪夫不等式的证明骨架; 理解方差快捷公式、线性变换、协方差性质、 $|\rho| \leq 1$ 及等号条件的核心转换.
- **C 级:** 知道期望线性性的离散与连续严格证明、 $D(X) = 0$ 的严格等价证明, 以及相关系数有界性的高层不等式细节, 本阶段不要求完整复现.

自测题

练习 4.35

设离散型随机变量 X 的分布律为

$$P(X = -1) = \frac{1}{4}, \quad P(X = 0) = \frac{1}{2}, \quad P(X = 2) = \frac{1}{4}.$$

求 $E(X), E(X^2), D(X)$, 并用 $D(3X - 2) = 9D(X)$ 检查 $3X - 2$ 的方差.

练习 4.36

设 $E(X) = 2, D(X) = 3$. 不假设具体分布, 利用切比雪夫不等式给出

$$P(-1 < X < 5)$$

的下界, 并说明为什么这只是一个普适下界.

练习 4.37

设 $D(X) = 4, D(Y) = 9, \text{Cov}(X, Y) = -3$. 求 $D(2X - Y)$ 与 ρ_{XY} , 并判断二者线性变化方向.

练习 4.38

构造或复述一个“不相关但不独立”的例子, 并分别指出“不相关”和“不独立”的判定依据. 若再补充 (X, Y) 联合正态, 结论如何变化?

注: [章末自测清单] 合上讲义后, 检查自己能否做到:

1. 看到分布就能选择求期望和方差的公式, 并说明存在条件;
2. 求函数期望时不必先求新变量分布;
3. 写线性组合方差时不遗漏协方差项;
4. 把切比雪夫事件整理到以期望为中心;
5. 解释协方差与相关系数的区别;

6. 用反例说明“不相关不等于独立”；
7. 说清二维正态中零相关与独立等价的前提.

若其中任一项不能独立完成, 应回到对应小节重做例题, 而不是只背结论.

第五章 大数定律和中心极限定理

本章导读

前四章主要研究一个或少数随机变量. 本章把视角转向大量随机变量: 重复试验越来越多时, 随机性会不会显现出稳定规律?

本章的两个核心问题是:

1. **大数定律:** 大量观测的算术平均会稳定到哪里?
2. **中心极限定理:** 大量独立随机变量的和或平均值, 近似服从什么分布?

二者都研究“大量重复”, 但回答的问题不同:

大数定律: 稳定值

中心极限定理: 近似分布.

本章依赖第四章的期望、方差、切比雪夫不等式和正态标准化; 下一章讨论样本统计量时, 会反复使用这里的结论.

5.1 依概率收敛

5.1.1 为什么需要一种新的收敛

普通数列 $x_n \rightarrow a$ 表示: 当 n 足够大时, x_n 必然落在 a 附近. 但随机变量 X_n 的取值具有随机性, 不能要求每次取值都落在附近. 我们能要求的是: n 越大, X_n 落在 a 附近的概率越接近 1.

定义 5.1 (依概率收敛)

设 $\{X_n\}$ 为随机变量序列, a 为常数. 若对任意 $\varepsilon > 0$, 都有

$$\lim_{n \rightarrow \infty} P(|X_n - a| < \varepsilon) = 1,$$

则称 X_n 依概率收敛于 a , 记为

$$X_n \xrightarrow{P} a.$$

等价地,

$$\lim_{n \rightarrow \infty} P(|X_n - a| \geq \varepsilon) = 0.$$

注: 定义中有三个不能漏掉的词: 任意 $\varepsilon > 0$ 、概率的**极限**、极限等于 1 或 0. 依概率收敛不表示从某一项起每次都落入邻域, 也不表示误差随 n 单调减小.

例题 5.2 (从定义识别依概率收敛)

设

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

证明 $X_n \xrightarrow{P} 0$.

任取 $\varepsilon > 0$. 当 $n > \varepsilon$ 时,

$$P(|X_n - 0| \geq \varepsilon) = P(X_n = n) = \frac{1}{n} \rightarrow 0.$$

故 $X_n \xrightarrow{P} 0$.

结果检查: 虽然 X_n 偶尔会取很大的值 n , 但这种异常取值的概率只有 $1/n$, 最终趋于零. 这正体现了“允许偶尔偏离, 但偏离概率消失”.

5.1.2 连续函数下的运算性质

定理 5.3 (依概率收敛的连续映射性质)

若 $X_n \xrightarrow{P} a$, 且函数 g 在 a 处连续, 则

$$g(X_n) \xrightarrow{P} g(a).$$

若 $X_n \xrightarrow{P} a$ 、 $Y_n \xrightarrow{P} b$, 且二元函数 g 在 (a, b) 处连续, 则

$$g(X_n, Y_n) \xrightarrow{P} g(a, b).$$

常用特例包括

$$X_n + Y_n \xrightarrow{P} a + b, \quad X_n Y_n \xrightarrow{P} ab,$$

以及当 $b \neq 0$ 时

$$\frac{X_n}{Y_n} \xrightarrow{P} \frac{a}{b}.$$

注: 使用函数结论前必须检查连续性. 例如倒数函数 $g(x) = 1/x$ 在 0 处不连续, 所以由 $X_n \xrightarrow{P} 0$ 不能直接推出 $1/X_n$ 的收敛结论.

小结

依概率收敛表示随机变量偏离目标常数至少 ε 的概率趋于零. 判断时固定从定义的两等价形式出发; 进行函数变换时, 还要检查函数在极限点是否连续. 后面所有大数定律的结论, 都是依概率收敛.

5.2 大数定律

5.2.1 共同主线: 算术平均稳定到期望

令

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

大数定律研究的典型结论是

$$\bar{X}_n - E(\bar{X}_n) \xrightarrow{P} 0,$$

或在同分布情形下

$$\bar{X}_n \xrightarrow{P} \mu.$$

直观上, 大量观测取平均会削弱个别异常值的影响, 使平均值稳定在总体期望附近.

5.2.2 切比雪夫大数定律

定理 5.4 (切比雪夫大数定律)

设随机变量序列 $X_1, X_2, \dots$ 满足:

1. $X_1, X_2, \dots$ 两两不相关;
2. 各随机变量的期望和方差存在;
3. 方差一致有界, 即存在常数 $C > 0$, 使

$$D(X_i) \leq C, \quad i = 1, 2, \dots.$$

则对任意 $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \varepsilon\right) = 1.$$

即

$$\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i) \xrightarrow{P} 0.$$

特别地, 若 $E(X_i) = \mu$, 则

$$\bar{X}_n \xrightarrow{P} \mu.$$

注: [怎样由切比雪夫不等式得到大数定律] 这是本章需要理解的 B 级证明骨架. 令

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

两两不相关使和的方差可以相加, 因此

$$D(Y_n) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) \leq \frac{nC}{n^2} = \frac{C}{n}.$$

由切比雪夫不等式,

$$P(|Y_n - E(Y_n)| \geq \varepsilon) \leq \frac{D(Y_n)}{\varepsilon^2} \leq \frac{C}{n\varepsilon^2} \rightarrow 0.$$

这恰好是 $Y_n - E(Y_n) \xrightarrow{P} 0$ 的定义.

真正起作用的链条是

$$\boxed{\text{两两不相关}} \implies \boxed{\text{平均值方差可相加}} \implies \boxed{D(Y_n) \rightarrow 0} \implies \boxed{\text{依概率收敛}}.$$

5.2.3 辛钦大数定律

定理 5.5 (辛钦大数定律)

设 $X_1, X_2, \dots$ 独立同分布, 且

$$E|X_1| < \infty, \quad E(X_1) = \mu.$$

则

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu.$$

注: “期望存在”严格地说是 $E|X_1| < \infty$. 辛钦大数定律不要求方差存在, 因此适用于某些方差无穷但期望有限的分布. 其严格证明属于 C 级, **本阶段不要求复现**; 必须准确记住条件和结论, 并会逐项验证.

5.2.4 两种大数定律的辨析

比较项	切比雪夫大数定律	辛钦大数定律
变量关系	两两不相关即可	独立同分布
分布要求	可以不同分布	必须同分布
矩条件	方差存在且一致有界	只要求绝对一阶矩有限
结论	平均值减其期望依概率趋于零	平均值依概率趋于共同期望
证明要求	理解切比雪夫不等式推导	严格证明本阶段不要求复现

注: [方法选择] 若题目直接给出“独立同分布”且只需研究算术平均, 通常优先用辛钦大数定律, 因为只需检查期望存在. 若变量不同分布, 但能验证两两不相关且方差一致有界, 则使用切比雪夫大数定律.

5.2.5 伯努利大数定律与频率稳定性

定理 5.6 (伯努利大数定律)

在 n 次独立重复试验中, 事件 A 每次发生的概率为 p , 发生次数为 N_n . 则对任意 $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{N_n}{n} - p\right| < \varepsilon\right) = 1,$$

即

$$\frac{N_n}{n} \xrightarrow{P} p.$$

注: 伯努利大数定律把第一章“频率稳定于概率”的直观说法写成了严格的概率极限. 它不保证试验次数每增加一次, 频率就更接近 p ; 它保证偏离 p 超过固定误差的概率趋于零.

证明: **证明目标:** 说明事件频率依概率收敛于事件概率.

出发点: 把每次试验是否发生 A 写成 0-1 随机变量, 并使用辛钦大数定律.

关键转换: 令

$$I_i = \begin{cases} 1, & \text{第 } i \text{ 次试验中 } A \text{ 发生,} \\ 0, & \text{第 } i \text{ 次试验中 } A \text{ 不发生.} \end{cases}$$

则频率就是这些指示变量的算术平均.

正式推导: $I_1, I_2, \dots$ 独立同分布, 且

$$E(I_i) = p, \quad N_n = \sum_{i=1}^n I_i.$$

由辛钦大数定律,

$$\frac{N_n}{n} = \frac{1}{n} \sum_{i=1}^n I_i \xrightarrow{P} p.$$

证明小结: 频率稳定性本质上是 0-1 随机变量样本均值的稳定性. 可复用的方法是先把“次数”写成指示变量之和, 再调用大数定律. ■

5.2.6 用大数定律处理算术平均

例题 5.7 (平方的算术平均)

设 $X_1, X_2, \dots$ 相互独立, 且均服从参数为 2 的指数分布. 求

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i^2$$

的依概率极限.

方法选择: 研究的是 X_i^2 的算术平均, 所以应验证 X_i^2 , 而不是只验证 X_i 的条件.

因为 X_i 独立同分布, 所以 X_i^2 也独立同分布. 又由指数分布的数字特征,

$$E(X_i) = \frac{1}{2}, \quad D(X_i) = \frac{1}{4},$$

故

$$E(X_i^2) = D(X_i) + [E(X_i)]^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < \infty.$$

由辛钦大数定律,

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \frac{1}{2}.$$

关键转折: “求谁的算术平均, 就验证谁的条件”. 这里的基本变量是 X_i^2 .

例题 5.8 (由密度确定依概率极限)

设 $X_1, X_2, \dots$ 独立同分布, 密度为

$$f(x) = \begin{cases} xe^{-x^2/2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

判断是否存在常数 a , 使

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} a,$$

若存在, 求 a .

仍然验证 X_i^2 . 它们独立同分布, 且

$$E(X_i^2) = \int_0^{+\infty} x^2 \cdot xe^{-x^2/2} dx.$$

令 $u = x^2/2$, 则 $x dx = du$, $x^2 = 2u$, 故

$$E(X_i^2) = \int_0^{+\infty} 2ue^{-u} du = 2.$$

由辛钦大数定律,

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} 2.$$

因此所求常数存在, 且 $a = 2$.

结果检查: 题目问“是否存在”, 不能只写数值; 必须先验证定律条件, 再明确回答存在性.

小结

大数定律研究算术平均的稳定性. 切比雪夫大数定律适合两两不相关、方差一致有界的序列; 辛钦大数定律适合独立同分布且绝对一阶矩有限的序列; 伯努利大数定律说明频率依概率收敛于概率. 做题时先识别被平均的变量, 再逐项验证相应条件.

5.3 独立同分布中心极限定理

5.3.1 从“稳定值”走向“近似分布”

大数定律告诉我们 $\bar{X}_n$ 会靠近 μ , 但没有描述它在 μ 附近怎样波动. 中心极限定理进一步回答: 经过适当中心化和标准化后, 这种波动的分布趋近标准正态分布.

比较项	大数定律	中心极限定理
主要对象	算术平均	和或算术平均的标准化形式
回答问题	最终稳定到哪个常数	最终近似哪种分布
典型结论	$\bar{X}_n \xrightarrow{P} \mu$	标准化后分布趋近 $N(0, 1)$
主要用途	解释频率、样本均值的稳定性	近似计算大量求和或平均的概率

5.3.2 列维-林德伯格中心极限定理

定理 5.9 (独立同分布中心极限定理)

设 $X_1, X_2, \dots$ 独立同分布, 且

$$E(X_i) = \mu, \quad D(X_i) = \sigma^2 > 0.$$

记

$$S_n = \sum_{i=1}^n X_i.$$

则对任意实数 x ,

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x),$$

其中

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

是标准正态分布函数.

定理常写成近似形式: 当 n 足够大时,

$$S_n \sim N(n\mu, n\sigma^2),$$

或者

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

这里的 $\sim$ 表示“近似服从”, 不是精确分布相等.

注: [为什么要中心化和标准化] 和 S_n 的期望为 $n\mu$, 方差为 $n\sigma^2$. 减去 $n\mu$ 是把中心移到 0; 除以 $\sigma\sqrt{n}$ 是把方差缩放到 1. 于是不同 n 、不同量纲的问题都被转化为同一个标准正态分布.

中心极限定理的严格证明通常使用特征函数、矩母函数或高等分析工具, 属于 C 级, 本阶段不要求复现. 现阶段必须会准确检查条件、完成标准化并计算近似概率.

5.3.3 样本均值形式

由

$$\bar{X}_n = \frac{S_n}{n}$$

可得

$$E(\bar{X}_n) = \mu, \quad D(\bar{X}_n) = \frac{\sigma^2}{n}.$$

因此当 n 足够大时,

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

且

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

注: 若原变量本身服从正态分布, 则独立正态变量的和与平均值**精确**服从正态分布, 不需要“大样本近似”; 若原分布不是正态, 上述正态结论来自中心极限定理, 只是近似.

5.3.4 中心极限定理的解题流程

面对应用题, 可按以下固定顺序处理:

1. **识别对象:** 题目是否研究大量个体的和或平均值;
2. **设随机变量:** 对谁求和, 就把第 i 个个体对应的量设为 X_i ;
3. **检查条件:** 说明 X_i 独立同分布, 且期望、正方差存在;
4. **确定参数:** 求 S_n 或 $\bar{X}_n$ 的期望与方差;
5. **标准化:** 减期望、除标准差;
6. **查表计算:** 转化为 $\Phi(x)$, 并利用 $\Phi(-x) = 1 - \Phi(x)$.

例题 5.10 (样本平均重量的近似概率)

某批棉花每包重量的期望为 100 kg, 标准差为 5 kg. 独立抽取 100 包, 求样本平均重量小于 99.5 kg 的近似概率. 已知 $\Phi(1) = 0.8413$.

设第 i 包棉花的重量为 X_i kg, $i = 1, 2, \dots, 100$. 则

$$E(X_i) = 100, \quad D(X_i) = 25.$$

由中心极限定理,

$$\bar{X} \sim N\left(100, \frac{25}{100}\right).$$

所以

$$\begin{aligned} P(\bar{X} < 99.5) &\approx P\left(\frac{\bar{X} - 100}{5/\sqrt{100}} < \frac{99.5 - 100}{0.5}\right) \\ &= \Phi(-1) = 1 - \Phi(1) = 0.1587. \end{aligned}$$

方法选择: 题目研究 100 个独立重量的平均值, 且只给出总体期望和方差, 正是中心极限定理的使用场景.

结果检查: 阈值 99.5 比均值低一个“样本均值标准差”, 左尾概率应小于 0.5, 结果合理. 题目中的总体包数若不影响抽样模型, 就是干扰信息.

例题 5.11 (寿命总和的近似概率)

某种电子设备的寿命服从均值为 100 小时的指数分布. 随机抽取 16 台, 寿命相互独立. 求寿命总和超过 1920 小时的近似概率. 已知 $\Phi(0.8) = 0.7881$.

设第 i 台设备寿命为 X_i . 指数分布的标准差等于均值, 因此

$$E(X_i) = 100, \quad D(X_i) = 100^2.$$

令 $S_{16} = \sum_{i=1}^{16} X_i$. 由中心极限定理,

$$S_{16} \sim N(16 \times 100, 16 \times 100^2).$$

于是

$$\begin{aligned} P(S_{16} > 1920) &\approx P\left(\frac{S_{16} - 1600}{400} > \frac{1920 - 1600}{400}\right) \\ &= 1 - \Phi(0.8) = 0.2119. \end{aligned}$$

近似性提醒: $n = 16$ 不算很大, 因此结果只是一种题设要求下的近似. 样本较小时, 中心极限定理的精度取决于原分布偏斜程度; 若题目能提供总和的精确分布, 应优先使用精确分布.

小结

独立同分布中心极限定理说明: 有限期望、正方差条件下, 大量随机变量之和经过“减 $n\mu$ 、除 $\sigma\sqrt{n}$ ”后趋近标准正态分布. 求平均值时相应地“减 μ 、除 $\sigma/\sqrt{n}$ ”. 应用中先识别大量求和或平均, 再设变量、验条件、定参数、标准化.

5.4 二项分布的正态近似

5.4.1 棣莫弗–拉普拉斯定理

二项随机变量本身就是 n 个独立 0–1 随机变量之和, 因此中心极限定理可直接给出它的正态近似.

定理 5.12 (棣莫弗–拉普拉斯定理)

设 $X_n \sim B(n, p)$, 其中 $0 < p < 1$ 固定. 则对任意实数 x ,

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

因此当 n 足够大时,

$$X_n \sim N(np, np(1-p)).$$

注: 其严格证明属于 C 级, 本阶段不要求复现. 做题时应能解释: X_n 是 n 个独立同分布的伯努利变量之和, 所以它是独立同分布中心极限定理的特例.

5.4.2 什么时候使用正态近似

常用经验条件是

$$np \geq 5, \quad n(1-p) \geq 5.$$

这不是定理中的精确分界线, 而是判断二项分布是否过度偏斜的经验规则. 两个量都足够大时, 正态近似通常较合理.

近似方法	典型条件	近似分布
泊松近似	n 大、 p 小、 np 适中	$P(np)$
正态近似	np 与 $n(1-p)$ 都不小	$N(np, np(1-p))$

注: [连续性修正] 二项分布是离散型, 正态分布是连续型. 若课程要求提高近似精度, 可把整数边界向外移动 0.5:

$$P(X \geq k) \approx P(Y > k - 0.5), \quad P(X \leq k) \approx P(Y < k + 0.5),$$

其中 $Y \sim N(np, np(1-p))$. 若题目或课堂示例明确不采用连续性修正, 应按其口径计算, 并说明所用近似.

例题 5.13 (考试通过人数)

某公司有 200 名员工参加资格考试. 根据往年经验, 每人通过的概率为 0.8, 并假定各人是否通过相互独立. 求至少 150 人通过的近似概率. 已知

$$\Phi\left(\frac{5\sqrt{2}}{4}\right) = 0.9616.$$

设通过人数为 X , 则

$$X \sim B(200, 0.8),$$

且

$$E(X) = 160, \quad D(X) = 200 \times 0.8 \times 0.2 = 32.$$

因为 $np = 160$ 、 $n(1 - p) = 40$ 均大于 5, 可用正态近似. 按题目给出的查表值, 不作连续性修正:

$$\begin{aligned} P(X \geq 150) &\approx P\left(\frac{X - 160}{\sqrt{32}} \geq \frac{150 - 160}{\sqrt{32}}\right) \\ &= P\left(Z \geq -\frac{5\sqrt{2}}{4}\right) = \Phi\left(\frac{5\sqrt{2}}{4}\right) = 0.9616. \end{aligned}$$

关键转折: “至少通过人数”先识别为二项分布; 由于 n 大且成功、失败的期望次数都不小, 再改用正态近似.

连续性修正: 若按更精细的口径, 应把 $X \geq 150$ 的边界改为 149.5, 再标准化. 两种结果会略有差异, 作答时保持口径一致.

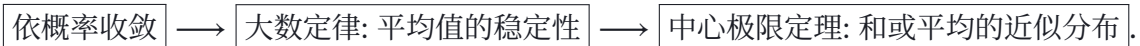
小结

棣莫弗-拉普拉斯定理把大样本二项分布近似为 $N(np, np(1 - p))$. 使用前确认二项模型, 并检查 np 与 $n(1 - p)$ 是否足够大; 随后减去 np 、除以 $\sqrt{np(1 - p)}$. 连续性修正用于改善离散分布到连续分布的近似精度.

5.5 本章总结

知识结构

本章可以按“先稳定、后波动”组织:



其中大数定律包括切比雪夫、辛钦和伯努利三种常用形式; 中心极限定理包括一般独立同分布形式及其二项分布特例.

核心结论与使用条件

结论	核心形式	使用前检查
依概率收敛	$P(X_n - a \geq \varepsilon) \rightarrow 0$	对任意 $\varepsilon > 0$
切比雪夫大数定律	平均值减其期望 $\xrightarrow{P} 0$	两两不相关、方差存在且一致有界
辛钦大数定律	$\bar{X}_n \xrightarrow{P} \mu$	独立同分布、 $E X_1 < \infty$
伯努利大数定律	$N_n/n \xrightarrow{P} p$	独立重复试验、成功概率固定
独立同分布中心极限定理	$(S_n - n\mu)/(\sigma\sqrt{n})$ 近似 $N(0, 1)$	独立同分布、有限期望、正方差
棣莫弗-拉普拉斯	$(X - np)/\sqrt{np(1-p)}$ 近似 $N(0, 1)$	$X \sim B(n, p)$, n 足够大

方法选择

1. 看到依概率极限: 先判断是否为算术平均, 再验证大数定律条件.
2. 看到独立同分布平均: 优先尝试辛钦大数定律, 只需再检查被平均变量的期望存在.
3. 看到不同分布序列: 若两两不相关且方差一致有界, 使用切比雪夫大数定律.
4. 看到频率: 把发生次数写成指示变量之和, 调用伯努利大数定律.
5. 看到大量和或平均的概率: 设第 i 个个体为 X_i , 验中心极限定理条件并标准化.
6. 看到大样本二项分布: 检查 np 与 $n(1-p)$, 再用正态近似; 按课程口径决定是否连续性修正.

重点辨析

容易混淆的概念	正确理解	常见误判
依概率收敛与普 通收敛	偏离概率趋于零, 仍允许偶然偏 离	认为从某项起每次都接近极限
大数定律与中心 极限定理	前者给稳定值, 后者给近似分布	看到“大量”就随意套其中一个
切比雪夫与辛钦	前者弱化独立同分布, 后者弱化 方差要求	混记三个条件和两个条件
原变量与被平均 变量	平均 $g(X_i)$ 就验证 $g(X_i)$ 的条件	只验证 X_i , 不检查变换后的矩
精确正态与近似 正态	正态总体的线性组合精确正态; 一般总体靠中心极限定理近似	把近似号写成严格分布相等
标准差与方差	标准化必须除标准差	直接除以 $n\sigma^2$ 或 σ^2/n

易错点清单

- 依概率收敛定义漏写“对任意 $\varepsilon > 0$ ”;
- 把“偏离概率趋于零”写成“随机变量数值必然趋于零”;
- 使用大数定律时不验证独立性、同分布性或矩条件;
- 研究 X_i^2 、 $g(X_i)$ 的平均, 却只检查 X_i 的条件;
- 切比雪夫大数定律忘记方差一致有界;
- 中心极限定理把和的方差写成 $n^2\sigma^2$;
- 平均值标准化时把标准差写成 σ/n , 正确值应为 $\sigma/\sqrt{n}$;
- 把中心极限定理的正态近似当成精确分布;
- 二项分布正态近似时忘记 $np(1-p)$ 是方差, 标准化分母还要开平方;
- 左尾、右尾转换时误用 $\Phi(-x) = 1 - \Phi(x)$.

证明掌握清单

- **A 级:** 能准确写出切比雪夫大数定律、辛钦大数定律的条件与结论; 能验证条件并完成依概率收敛或中心极限定理计算.
- **B 级:** 能解释切比雪夫大数定律怎样由切比雪夫不等式推出; 能说明大数定律为何支持频率稳定性和用样本矩估计总体矩; 能解释中心化与标准化的作用.
- **C 级:** 知道辛钦大数定律、独立同分布中心极限定理和棣莫弗—拉普拉斯定理的严格证明结论与用途; 使用特征函数、矩母函数或高等分析的证明本阶段不要求复现.

自测题

练习 5.14

设

$$P(X_n = 1) = 1 - \frac{1}{n^2}, \quad P(X_n = n) = \frac{1}{n^2}.$$

从定义证明 $X_n \xrightarrow{P} 1$, 并判断能否由连续映射性质推出 $X_n^2 \xrightarrow{P} 1$.

练习 5.15

设 $X_1, X_2, \dots$ 独立同分布, $E(X_1) = 2, D(X_1) = 3$. 分别写出

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

的依概率极限, 以及其大样本近似正态分布. 说明两个结论分别来自哪个定理.

练习 5.16

设 $X_1, X_2, \dots$ 独立同分布, $E(X_1^4) < \infty$. 证明

$$\frac{1}{n} \sum_{i=1}^n X_i^4 \xrightarrow{P} E(X_1^4),$$

并明确指出应验证哪个随机变量序列的条件.

练习 5.17

某商品单件重量的期望为 500 g, 标准差为 20 g. 独立抽取 100 件, 利用中心极限定理表示样本平均重量落在 496 g 到 504 g 之间的近似概率, 并把答案写成 Φ 的形式.

练习 5.18

设 $X \sim B(400, 0.02)$. 比较泊松近似与正态近似的适用条件, 并判断本题优先选择哪一种近似. 若改为 $X \sim B(400, 0.5)$, 结论如何变化?

注: [章末自测清单] 合上讲义后, 检查自己能否做到:

1. 用两种等价形式写出依概率收敛定义;
2. 不看讲义写出切比雪夫与辛钦大数定律的条件和结论;
3. 说明“频率稳定于概率”的严格含义;
4. 面对 $g(X_i)$ 的平均, 正确验证 $g(X_i)$ 的条件;
5. 区分大数定律与中心极限定理分别回答什么问题;
6. 写出和与平均值的中心极限定理标准化式;
7. 识别二项分布正态近似并说明连续性修正的作用.

若有任一项不能独立完成, 应回到相应小节重做例题, 而不是只背最终公式.

第六章 数理统计的基本概念

本章导读

前五章通常从一个已经确定的分布出发, 计算概率或数字特征. 本章把方向反过来: 总体分布含有未知信息时, 怎样从有限个观测值推断总体? 这就是数理统计的起点.

本章的知识链是



前三个框说明“用什么数据、怎样加工数据”, 最后一个框说明“加工结果具有怎样的随机规律”. 第七至第九章的参数估计和假设检验, 都建立在这条链上.

6.1 总体与样本

6.1.1 总体不是一堆编号, 而是所研究指标的分布

假设我们关心一批灯泡的寿命. 每只灯泡是一个**个体**, 寿命是个体上的数量指标. 若从总体中随机抽取一只灯泡, 用随机变量 X 表示其寿命, 那么 X 的分布就刻画了总体.

定义 6.1 (总体、个体与总体分布)

研究对象的全体称为**总体**, 组成总体的每个基本单位称为**个体**. 在数理统计中, 通常用个体上某个数量指标的分布来代表总体, 也把服从该分布的随机变量 X 简称为总体.

总体分布可能完全已知, 也可能只知道分布族而不知道参数. 例如

$$X \sim N(\mu, \sigma^2)$$

表示已知总体属于正态分布族; 若 μ, σ^2 未知, 统计问题就是利用样本获得关于它们的信息.

注: [模型中的“已知”与“未知”] “总体是正态总体”不等于“总体分布完全已知”. 前者只说明分布类型; 只有当 μ 和 σ^2 也已知时, 分布才被完全确定.

6.1.2 简单随机样本

不可能观察整个总体时, 就从中抽取一部分个体. 为使样本能够代表总体, 抽样方式必须受到条件约束.

定义 6.2 (简单随机样本)

设总体的分布函数为 $F(x)$. 若随机变量 $X_1, X_2, \dots, X_n$ 满足:

1. $X_1, \dots, X_n$ 相互独立;
2. 每个 X_i 都与总体 X 同分布, 分布函数均为 $F(x)$,

则称 $X_1, \dots, X_n$ 是来自总体 X 的容量为 n 的**简单随机样本**, 简称**样本**.

“同分布”保证每次抽取面对同一个总体规律; “相互独立”保证一次抽取的结果不改变其他抽取的概率规律. 本讲义后续所说的样本, 若无特别说明, 均指简单随机样本.

6.1.3 样本的二重性

抽样前, $X_1, \dots, X_n$ 的值尚未确定, 它们是随机变量; 抽样后得到

$$x_1, x_2, \dots, x_n,$$

这些是确定的观测值. 二者不能混用.

阶段	记号	含义
观测前	$X_1, \dots, X_n$	随机变量, 可讨论分布、期望和方差
观测后	$x_1, \dots, x_n$	已经得到的数, 可进行具体计算

例题 6.3 (区分样本与样本值)

随机抽取三只灯泡并记录寿命. 抽取前用 X_1, X_2, X_3 表示三次寿命, 它们构成随机样本; 若实际得到 1010, 980, 1050 小时, 这三个数是该样本的一组样本值.

为什么要区分：“样本均值服从什么分布”讨论的是随机变量的函数；“样本均值是多少”则把实际数字代入计算. 前者属于理论推断, 后者属于数据处理.

小结

总体用所研究指标的概率分布刻画; 简单随机样本必须同时满足独立和同分布. 使用样本前, 先辨明分布类型与参数哪些已知, 再辨明当前记号表示随机样本还是已经观测到的样本值.

6.2 经验分布函数

6.2.1 用样本中“不超过给定值的比例”模仿总体概率

总体分布函数是 $F(x) = P(X \leq x)$. 若 F 未知, 自然可以用样本中不超过 x 的观测所占比例来估计它.

定义 6.4 (经验分布函数)

设 $X_1, \dots, X_n$ 为样本, 定义

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}},$$

其中示性函数在事件 $\{X_i \leq x\}$ 发生时取 1, 否则取 0. 观测到样本值后, $F_n(x)$ 就是样本中不超过 x 的数据比例.

例题 6.5 (从频数写出经验分布函数)

一组容量为 10 的样本值为

1, 2, 4, 3, 4, 2, 3, 4, 4, 2.

不同取值 1, 2, 3, 4 的频数依次为 1, 3, 2, 4, 故

$$F_{10}(x) = \begin{cases} 0, & x < 1, \\ 0.1, & 1 \leq x < 2, \\ 0.4, & 2 \leq x < 3, \\ 0.6, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

方法选择: 先排序并统计不同取值的频数, 再逐段累加相对频数. **结果检查:** 经验分布函数必须右连续、单调不减, 且两端极限分别为 0 和 1.

定理 6.6 (格里文科定理)

设 $F_n(x)$ 是来自总体分布函数 $F(x)$ 的经验分布函数, 则当 $n \rightarrow \infty$ 时, $F_n(x)$ 以概率 1 一致收敛于 $F(x)$, 即

$$P\left(\lim_{n \rightarrow \infty} \sup_x |F_n(x) - F(x)| = 0\right) = 1.$$

注: [为什么经验分布能代表总体] 对固定的 x , 每个 $\mathbf{1}_{\{X_i \leq x\}}$ 都是参数为 $F(x)$ 的伯努利随机变量, $F_n(x)$ 正是它们的样本均值. 大数定律说明它会稳定到 $F(x)$. 格里文科定理进一步把“固定一个 x ”加强为“对所有 x 同时控制”. 严格证明涉及本阶段尚未学习的工具, **本阶段不要求复现**; 应会构造和检查经验分布函数.

小结

经验分布函数用样本比例估计总体分布函数, 是一条从数据到分布的直接通道. 构造时先排序、计频数、再累加; 检查时看它是否为右连续的阶梯函数, 并满足单调性和两端极限.

6.3 统计量与常用样本统计量

6.3.1 统计量: 只加工样本, 不偷用未知信息

定义 6.7 (统计量与统计量的观测值)

设 $X_1, \dots, X_n$ 为样本. 若函数

$$T = T(X_1, \dots, X_n)$$

不含任何未知参数, 则称 T 为**统计量**. 把样本值代入后得到的数 $T(x_1, \dots, x_n)$, 称为统计量的**观测值**或**统计值**.

统计量仍是随机变量, 因此有自己的分布、期望和方差. 定义中禁止的是**未知参数**, 而不是一切常数: 已知参数可以出现在统计量中.

例题 6.8 (判断是否为统计量)

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, 且 μ, σ^2 均未知. 则 $\bar{X}$ 、 $\max_i X_i$ 和 $\sum X_i^2$ 都是统计量; $\bar{X} - \mu$ 与 S^2/σ^2 含有未知参数, 不是统计量.

判断方法: 先列出题目中的未知参数, 再逐个检查表达式中是否出现它们. 不要因为一个表达式具有重要分布就误称它为统计量.

6.3.2 五类常用统计量

定义 6.9 (样本均值、样本方差与样本矩)

设 $X_1, \dots, X_n$ 为样本, 定义

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{样本均值,}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{样本方差,}$$

$$S = \sqrt{S^2} \quad \text{样本标准差,}$$

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad k \text{ 阶样本原点矩,}$$

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad k \text{ 阶样本中心矩.}$$

特别地, $A_1 = \bar{X}$, 而

$$B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2.$$

样本均值描述数据的中心位置; 样本方差和标准差描述离散程度; 样本矩则与总体矩对应, 是后面矩估计法的基础.

定理 6.10 (样本方差的计算恒等式)

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

因而

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right).$$

证明: 证明目标: 把“离均差平方和”改写为平方和与样本均值的组合.

出发点: 使用 $\sum_{i=1}^n X_i = n\bar{X}$.

关键转换: 展开平方, 再把与下标 i 无关的 $\bar{X}$ 提出求和.

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2. \end{aligned}$$

证明小结: 真正起作用的是样本均值定义. 这个恒等式既便于手算, 也是在证明 $E(S^2) = \sigma^2$ 时消去交叉项的关键工具. ■

小结

统计量是只由样本和已知量构成的函数; 代入样本值后才得到统计值. 遇到常用统计量, 要分清 S^2 的分母是 $n - 1$, 而二阶样本中心矩 B_2 的分母是 n .

6.4 统计量的数字特征

设总体 X 的期望和方差存在, 并记

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2.$$

定理 6.11 (样本均值的期望与方差)

对简单随机样本 $X_1, \dots, X_n$,

$$E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

证明: 证明目标: 求样本均值的期望和方差.

出发点: 样本同分布给出 $E(X_i) = \mu$ 、 $\text{Var}(X_i) = \sigma^2$; 样本独立使和的方差可以直接相加.

关键转换: 把常数 $1/n$ 提到期望或方差之外.

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu, \\ \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}. \end{aligned}$$

证明小结: 期望结论只需同分布, 方差相加还用到了独立性. 样本量扩大时, 样本均值的中心不变, 而波动按 $1/n$ 缩小. ■

定理 6.12 (样本方差的无偏性)

若总体方差为 σ^2 , 则

$$E(S^2) = \sigma^2.$$

因此用分母 $n - 1$ 定义的样本方差是总体方差的无偏估计.

证明: 证明目标: 证明离均差平方和除以 $n - 1$ 后, 期望恰为 σ^2 .

出发点: 使用恒等式

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

关键转换: 分别求右侧两项的期望; 第一项含 n 个总体方差, 第二项含样本均值的方差.

$$\begin{aligned} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= \sum_{i=1}^n E[(X_i - \mu)^2] - nE[(\bar{X} - \mu)^2] \\ &= n\sigma^2 - n\text{Var}(\bar{X}) \\ &= n\sigma^2 - n \cdot \frac{\sigma^2}{n} \\ &= (n-1)\sigma^2. \end{aligned}$$

两边除以 $n-1$, 得到 $E(S^2) = \sigma^2$.

证明小结: 估计均值 μ 消耗了一个自由度, 所以离均差平方和的期望只有 $(n-1)\sigma^2$. 若仍除以 n , 所得 B_2 的期望为 $\frac{n-1}{n}\sigma^2$, 会系统性偏小. ■

例题 6.13 (先识别结构, 再求期望)

若总体 $X \sim B(m, p)$, 求

$$E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right].$$

二项分布的方差为 $mp(1-p)$. 由上一证明中的中间结论, 直接得到

$$E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = (n-1)mp(1-p).$$

方法选择: 看到离均差平方和, 应先联系 $(n-1)S^2$, 不必逐项展开. **结果检查:** 当 $n=1$ 时离均差恒为零, 公式也给出零.

小结

样本均值满足 $E(\bar{X}) = \mu$ 、 $\text{Var}(\bar{X}) = \sigma^2/n$; 样本方差满足 $E(S^2) = \sigma^2$. 使用这些结论前, 先检查是否为简单随机样本, 以及所需的总体矩是否存在.

6.5 三大抽样分布与上分位点

6.5.1 抽样分布与上分位点

定义 6.14 (抽样分布)

统计量的概率分布称为该统计量的**抽样分布**. 它描述重复抽样时统计量怎样波动, 是进行区间估计和假设检验的概率基础.

定义 6.15 (上 α 分位点)

设连续型随机变量 Y 的密度为 $f(y)$. 若数 y_α 满足

$$P(Y > y_\alpha) = \int_{y_\alpha}^{\infty} f(y) dy = \alpha,$$

则称 y_α 为 Y 的上 α 分位点.

“上”表示右尾. 看到上 α 分位点, 应立刻画出右侧面积为 α 的示意图, 避免把 $P(Y \leq y_\alpha) = \alpha$ 写反.

6.5.2 卡方分布: 独立标准正态的平方和

定义 6.16 (χ^2 分布)

若 $Z_1, \dots, Z_k$ 相互独立且均服从 $N(0, 1)$, 则

$$V = \sum_{i=1}^k Z_i^2$$

服从自由度为 k 的卡方分布, 记为 $V \sim \chi^2(k)$.

常用性质为

$$E(V) = k, \quad \text{Var}(V) = 2k.$$

若 $V_1 \sim \chi^2(k_1)$ 、 $V_2 \sim \chi^2(k_2)$ 且相互独立, 则

$$V_1 + V_2 \sim \chi^2(k_1 + k_2).$$

自由度为 k 的卡方分布上 α 分位点记为 $\chi_\alpha^2(k)$.

例题 6.17 (线性组合先标准化, 再平方求和)

设 $X_1, \dots, X_5$ 相互独立且均服从 $N(0, 1)$, 令

$$Y = (X_1 + X_2 + X_3)^2 + (X_4 - \sqrt{2}X_5)^2.$$

求常数 c , 使 $cY \sim \chi^2(2)$.

两个括号均服从 $N(0, 3)$, 且因使用了互不重叠的独立变量而相互独立. 因此

$$\frac{X_1 + X_2 + X_3}{\sqrt{3}}, \quad \frac{X_4 - \sqrt{2}X_5}{\sqrt{3}}$$

是两个独立标准正态变量, 故 $Y/3 \sim \chi^2(2)$, 即 $c = 1/3$.

结果检查: 自由度等于独立标准正态平方项的个数, 而不是原始变量的总数.

6.5.3 学生分布: 标准正态除以独立卡方均方的平方根

定义 6.18 (t 分布)

若 $Z \sim N(0, 1)$ 、 $V \sim \chi^2(k)$, 且 Z 与 V 相互独立, 则

$$T = \frac{Z}{\sqrt{V/k}}$$

服从自由度为 k 的 t 分布, 记为 $T \sim t(k)$.

t 分布关于原点对称, 因此其上分位点满足

$$t_{1-\alpha}(k) = -t_{\alpha}(k).$$

当自由度增大时, $t(k)$ 逐渐接近 $N(0, 1)$; $t(1)$ 是标准柯西分布, 期望不存在. t 分布密度的严格推导属于本阶段不要求复现的深层理论证明.

6.5.4 F 分布: 两个独立卡方均方之比

定义 6.19 (F 分布)

若 $U \sim \chi^2(k_1)$ 、 $V \sim \chi^2(k_2)$, 且 U, V 相互独立, 则

$$F = \frac{U/k_1}{V/k_2}$$

服从自由度为 (k_1, k_2) 的 F 分布, 记为 $F \sim F(k_1, k_2)$.

定理 6.20 (t 与 F 的平方关系)

若 $T \sim t(k)$, 则

$$T^2 \sim F(1, k).$$

证明: 证明目标: 把 T^2 改写成两个独立卡方变量分别除以自由度后的比.

出发点: 由定义, $T = Z/\sqrt{V/k}$, 其中 $Z \sim N(0, 1)$ 、 $V \sim \chi^2(k)$ 且独立.

关键转换: 利用标准正态的平方服从 $\chi^2(1)$.

$$T^2 = \frac{Z^2/1}{V/k} \sim F(1, k).$$

证明小结: 平方把 t 分布分子中的标准正态变成自由度为 1 的卡方; 独立性由原定义保留. ■

定理 6.21 (F 分布的倒数与分位点)

若 $F \sim F(k_1, k_2)$, 则

$$\frac{1}{F} \sim F(k_2, k_1),$$

并且

$$F_{1-\alpha}(k_1, k_2) = \frac{1}{F_{\alpha}(k_2, k_1)}.$$

证明: 证明目标: 说明取倒数会交换两个自由度, 并推出相应的上分位点关系.

出发点: 写成 $F = (U/k_1)/(V/k_2)$, 其中两个卡方变量独立.

关键转换: 取倒数交换分子与分母.

$$\frac{1}{F} = \frac{V/k_2}{U/k_1} \sim F(k_2, k_1).$$

再由连续性,

$$P(F \leq F_{1-\alpha}(k_1, k_2)) = \alpha.$$

取倒数后不等号方向改变, 故

$$P\left(\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(k_1, k_2)}\right) = \alpha.$$

右端临界值应为 $F_{\alpha}(k_2, k_1)$, 整理即得结论.

证明小结: 倒数不仅交换自由度, 还会交换左右尾, 所以分位点下标由 $1 - \alpha$ 变为 α . ■

目标分布	典型结构	必查条件	自由度来源
χ^2	标准正态平方和	各项独立且已标准化	独立平方项数
t	标准正态除以卡方均方平方根	分子、分母独立	分母卡方自由度
F	两个卡方均方之比	两个卡方独立	分子、分母分别计数

注: [识别结构不等于只看外形] “平方和”不自动服从卡方, “正态除以根号平方和”也不自动服从 t . 每次都要完成三步: 先把正态变量标准化, 再核对自由度, 最后验证组成部分的独立性.

小结

三大抽样分布都由独立标准正态变量构造. 识别时按“标准化—独立性—自由度”检查; 查表时先确认使用的是上分位点, 并特别留意 F 分布取倒数会同时交换自由度和尾部.

6.6 一个正态总体的抽样分布

设 $X_1, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, $\bar{X}$ 和 S^2 分别为样本均值和样本方差.

定理 6.22 (正态总体的四个核心抽样结论)

1. 样本均值仍服从正态分布:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

2. $\bar{X}$ 与 S^2 相互独立.

3. 两个常用平方和满足

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1),$$

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n).$$

4. 用 S 代替未知的 σ 后,

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

注: [结论怎样串起来] 第一条来自独立正态变量的线性组合. 第三条中, 减去已知 μ 得到 n 个独立标准正态平方, 所以自由度为 n ; 减去样本均值后, 偏差满足

$$\sum_{i=1}^n (X_i - \bar{X}) = 0,$$

存在一个线性约束, 故只剩 $n-1$ 个自由方向. 第二条“正态样本均值与样本方差独立”的严格证明依赖正交变换, **本阶段不要求复现**; 但必须记住它依赖正态总体. 结合第二、三条就能按定义构造第四条.

证明: **证明目标:** 由标准正态与卡方的组合, 推出学生化统计量服从 $t(n-1)$.

出发点: 令

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

正态总体保证 $\bar{X}$ 与 S^2 独立, 所以 Z 与 V 独立.

关键转换: 按 t 分布定义组成 $Z/\sqrt{V/(n-1)}$.

$$\frac{Z}{\sqrt{V/(n-1)}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{S/\sigma} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

证明小结: 核心条件是总体正态和均值、方差的独立性. 若用 $\sum (X_i - \mu)^2$ 构造分母, 它一般不与 $\bar{X}$ 独立, 不能据此得到 $t(n)$. ■

例题 6.23 (样本均值的概率)

从 $N(5, 16)$ 中抽取容量为 36 的简单随机样本, 求

$$P(4 \leq \bar{X} \leq 6).$$

因为

$$\bar{X} \sim N\left(5, \frac{16}{36}\right),$$

其标准差为 $2/3$. 标准化得

$$P(4 \leq \bar{X} \leq 6) = P(-1.5 \leq Z \leq 1.5) = 2\Phi(1.5) - 1.$$

方法选择: 问题只涉及 $\bar{X}$ 且总体方差已知, 直接使用样本均值的正态分布. **结果检查:** 区间以总体均值 5 为中心, 标准化后应得到对称区间.

例题 6.24 (用卡方分布求统计量的方差)

设总体为 $N(0, 1)$, 令

$$T = \bar{X}^2 - \frac{S^2}{n}.$$

求 $E(T)$ 与 $\text{Var}(T)$.

由 $E(\bar{X}^2) = \text{Var}(\bar{X}) = 1/n$ 和 $E(S^2) = 1$, 得 $E(T) = 0$. 又因 $\bar{X}$ 与 S^2 独立, 平方后仍独立, 故

$$\text{Var}(T) = \text{Var}(\bar{X}^2) + \frac{1}{n^2} \text{Var}(S^2).$$

由 $n\bar{X}^2 \sim \chi^2(1)$ 和 $(n-1)S^2 \sim \chi^2(n-1)$,

$$\text{Var}(\bar{X}^2) = \frac{2}{n^2}, \quad \text{Var}(S^2) = \frac{2}{n-1}.$$

因此

$$\text{Var}(T) = \frac{2}{n^2} + \frac{2}{n^2(n-1)} = \frac{2}{n(n-1)}.$$

关键转折: 遇到“正态变量的平方”的方差, 先把整体改造成卡方变量, 比直接计算四阶矩更简洁.

小结

一个正态总体有四个必须熟练调用的结论: 样本均值的正态分布、均值与方差独立、两个平方和的卡方分布、学生化均值的 t 分布. 使用时先判断分母中减的是 μ 还是 $\bar{X}$, 再确定自由度是 n 还是 $n - 1$.

6.7 两个正态总体的抽样分布

设

$$X_1, \dots, X_{n_1} \sim N(\mu_1, \sigma_1^2), \quad Y_1, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2^2),$$

两组均为简单随机样本, 并且两组样本相互独立. 分别记样本均值和样本方差为 $\bar{X}, S_1^2$ 与 $\bar{Y}, S_2^2$.

定理 6.25 (两个样本均值之差)

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right),$$

从而

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1).$$

证明: 证明目标: 求两个独立正态样本均值之差的分布.

出发点:

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \quad \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right),$$

且两者独立.

关键转换: 独立正态变量的线性组合仍为正态变量; 作差时期望相减, 方差相加.

$$\begin{aligned} E(\bar{X} - \bar{Y}) &= \mu_1 - \mu_2, \\ \text{Var}(\bar{X} - \bar{Y}) &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}. \end{aligned}$$

由此得到第一式, 再按其均值和标准差标准化即得第二式.

证明小结: 作差不会让方差相减; 只有期望带负号, 方差仍相加. 两组样本相互独立是去掉协方差项的关键条件. ■

定理 6.26 (两个样本方差之比)

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

证明: **证明目标:** 把样本方差之比改造成两个独立卡方均方之比.

出发点:

$$U = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1), \quad V = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1),$$

且两组样本独立, 所以 U, V 独立.

关键转换: 分别除以各自自由度后作比.

$$\frac{U/(n_1 - 1)}{V/(n_2 - 1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

证明小结: 分子自由度来自第一组, 分母自由度来自第二组; 交换方差之比的次序时也要交换两个自由度. ■

定理 6.27 (等方差时的合并方差与两样本 t 统计量)

若 $\sigma_1^2 = \sigma_2^2 = \sigma^2$, 定义合并样本方差

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

则

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2).$$

注: [合并方差公式的构造骨架] 两个独立卡方变量具有可加性, 因此

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2).$$

另一方面, 均值差标准化后服从 $N(0, 1)$, 且均值部分与两个样本方差构成的 S_w^2 独立. 用前者除以后者除以自由度的平方根, 约去 σ , 便得到上述 t 统计量. 应掌握这条构造链; 完整的多元独立性证明本阶段不要求复现.

注: [三个公式怎样选择] 比较两个总体均值时, 若两个总体方差已知, 使用均值差的标准正态统计量; 若方差未知但已知相等, 使用合并方差构造的两样本 t 统计量. 比较两个总体方差时, 则从样本方差之比的 F 分布出发. 不要在未给出等方差条件时擅自使用 S_w^2 .

小结

两个正态总体的公式都可由一个总体的结论现推: 均值差用独立正态线性组合, 方差比用两个独立卡方均方之比, 等方差下的均值差再用合并方差学生化. 使用前必须检查两组样本相互独立, 以及合并方差所需的等方差条件.

本章总结

一、知识结构



二、核心公式及适用条件

结论	公式	关键条件
样本均值的数字特征	$E(\bar{X}) = \mu, \text{Var}(\bar{X}) = \sigma^2/n$	简单随机样本, 二阶矩存在
样本方差无偏性	$E(S^2) = \sigma^2$	简单随机样本, 二阶矩存在
一个正态总体	$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$	正态总体
一个正态总体	$(\bar{X} - \mu)/(S/\sqrt{n}) \sim t(n-1)$	正态总体
两个均值之差	标准化后服从 $N(0, 1)$	两组独立正态样本, 方差已知
两个方差之比	$(S_1^2/\sigma_1^2)/(S_2^2/\sigma_2^2) \sim F(n_1-1, n_2-1)$	两组独立正态样本
合并方差统计量	学生化后服从 $t(n_1 + n_2 - 2)$	两组独立正态样本且方差相等

三、常见题型与方法选择

1. **构造经验分布函数:** 排序、计频数、累计相对频数, 再检查右连续与两端极限.
2. **判断统计量:** 先列未知参数, 再检查表达式中是否含有它们.
3. **求期望或方差:** 优先识别 $\bar{X}$ 、 S^2 、离均差平方和等整体, 不要急于逐项展开.
4. **判断抽样分布:** 先标准化, 再验证独立性, 最后按平方项个数确定自由度.

5. **正态总体平方问题:** 看到标准化正态的平方, 联想到 $\chi^2(1)$; 看到两个卡方均方之比, 联想到 F .
6. **两个总体问题:** 先问目标是均值差还是方差比, 再问方差已知、未知相等还是未知且不说明相等.

四、重点辨析

概念	正确理解	常见误判
样本与样本值	前者是随机变量组, 后者是观测后的数	用小写数据讨论抽样分布
统计量	只要求不含未知参数	误以为不能含任何参数或常数
S^2 与 B_2	分母分别为 $n-1$ 与 n	把二者视为同一个量
n 与 $n-1$	减已知 μ 得 n ; 减估计量 $\bar{X}$ 得 $n-1$	只按平方项表面个数计数
独立与不相关	构造 t, F 必须验证独立性	只看表达式外形便套分布
已知方差与等方差	已知方差用正态; 未知但相等才用合并方差	未给等方差条件仍强行合并

五、易错点清单

- 把总体当作有限数据表, 而忘记统计模型研究的是指标分布.
- 只检查样本“同分布”, 漏掉简单随机样本还要求相互独立.
- 把含未知参数的枢轴量误称为统计量.
- 把 $\text{Var}(\bar{X})$ 写成 $\sigma^2/\sqrt{n}$; 方差是 σ^2/n , 标准差才是 $\sigma/\sqrt{n}$.
- 忘记 F 分布的两个自由度有先后顺序.
- 把 $\bar{X}$ 与 S^2 的独立性推广到任意总体; 该核心结论依赖正态总体.
- 作差时误把方差也相减; 独立变量之差的方差仍相加.

六、证明掌握清单

- **A 级, 应能独立复现:** 样本均值的期望与方差; 样本方差恒等式与无偏性; 由独立标准正态变量构造 χ^2, t, F ; $T^2 \sim F(1, n)$ 、 F 倒数和分位点关系; 两个独立正态样本均值之差与方差之比.
- **B 级, 应掌握证明骨架:** 样本方差卡方分布自由度为何为 $n-1$; 等方差时合并方差和两样本 t 统计量的构造; 正态抽样公式之间怎样通过标准化连接.
- **C 级, 准确记条件和用途:** 正态样本均值与样本方差独立的严格证明; 格里文科定理严格证

明; 三大抽样分布的密度推导; 正态抽样定理的完整多元几何证明. 这些内容**本阶段不要求复现**.

七、自测题

练习 6.28

设 $X_1, \dots, X_8$ 来自总体 $N(\mu, 4)$, μ 未知. 判断下列各式哪些是统计量:

$$\bar{X}, \quad \bar{X} - \mu, \quad \sum_{i=1}^8 (X_i - \bar{X})^2, \quad \frac{\sqrt{8}(\bar{X} - \mu)}{2}.$$

练习 6.29

一组样本值为 0, 1, 1, 2, 3. 写出其经验分布函数, 并检查右连续性、单调性和两端极限.

练习 6.30

设总体方差为 9, 样本容量为 25. 求 $E(\bar{X})$ 、 $\text{Var}(\bar{X})$ 与

$$E\left[\sum_{i=1}^{25} (X_i - \bar{X})^2\right].$$

练习 6.31

设 $X_1, \dots, X_{10}$ 来自 $N(0, 1)$. 分别判断

$$\sum_{i=1}^{10} X_i^2, \quad \sum_{i=1}^{10} (X_i - \bar{X})^2, \quad \frac{\sqrt{10}\bar{X}}{S}$$

服从什么分布, 并说明自由度来源.

练习 6.32

若 $T \sim t(12)$, 说明 T^2 和 $1/T^2$ 的分布; 若 $F \sim F(5, 8)$, 说明 $1/F$ 的分布, 并写出相应上分位点关系.

练习 6.33

两组相互独立的正态样本容量分别为 $n_1 = 10, n_2 = 15$. 说明在以下三种任务中应构造哪类分布: 总体方差均已知时比较均值; 总体方差未知但相等时比较均值; 比较两个总体方差.

完成自测后, 应能用一句话概括本章: 从总体抽取独立同分布样本, 用不含未知参数的统计量压缩数据, 再借助其抽样分布进行统计推断.

第七章 点估计与估计量的评价

本章导读

上一章解决了“怎样从总体取得样本、统计量具有怎样的分布”. 本章进一步追问: 总体分布的形式已知, 但参数未知时, 能否用样本给参数一个具体的估计?

本章分成两条主线:

怎样估: 矩估计、最大似然估计

估得怎样: 无偏性、有效性、一致性.

矩估计利用“样本矩接近总体矩”, 最大似然估计利用“已经出现的数据在所选参数下应尽可能容易出现”. 二者可能给出相同结果, 也可能不同; 得到估计量后, 还必须评价其质量.

7.1 点估计的基本概念

7.1.1 从未知参数到样本函数

设总体 X 的分布形式已知, 但含未知参数 θ ; $X_1, \dots, X_n$ 是来自该总体的简单随机样本. 我们的目标是构造一个不含未知参数的统计量, 用它近似 θ .

定义 7.1 (点估计、估计量与估计值)

若统计量

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$$

被用来估计未知参数 θ , 则称 $\hat{\theta}$ 为 θ 的**估计量**. 观测到样本值 $x_1, \dots, x_n$ 后, 所得数值

$$\hat{\theta}(x_1, \dots, x_n)$$

称为 θ 的**估计值**. 直接用一个数估计参数的方法称为**点估计**.

对象	典型记号	性质
未知参数	θ	固定但未知, 不是随机变量
估计量	$\hat{\theta}(X_1, \dots, X_n)$	随样本改变, 是统计量、随机变量
估计值	$\hat{\theta}(x_1, \dots, x_n)$	代入数据后得到的确定数值

帽子 “ $\widehat{}$ ” 表示估计, 不表示它一定等于真实参数. 点估计只给出一个数, 不直接说明误差可能有多大; 下一章的区间估计将补充不确定性范围.

小结

点估计的目标是把未知参数表示成样本的函数. 书写时必须区分固定的未知参数、随机的估计量和观测后的估计值; 后续评价的对象是估计量, 而不只是某次得到的数字.

7.2 矩估计法

7.2.1 思想来源: 样本矩逼近总体矩

设总体 k 阶原点矩存在, 记

$$\mu'_k = E(X^k), \quad A_k = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

由于 $X_1^k, \dots, X_n^k$ 独立同分布, 辛钦大数定律给出

$$A_k \xrightarrow{P} \mu'_k.$$

因此样本量较大时, 样本矩通常接近对应总体矩. 矩估计法把这种 “接近” 直接写成方程.

定义 7.2 (矩估计法)

设总体分布含 k 个未知参数 $\theta_1, \dots, \theta_k$. 选取 k 个存在的总体矩, 并令它们等于对应样本矩, 得到关于未知参数的方程组. 解出

$$\hat{\theta}_j = \widehat{\theta}_j(X_1, \dots, X_n), \quad j = 1, \dots, k,$$

称为各参数的矩估计量.

最常用的原点矩方程组为

$$\frac{1}{n} \sum_{i=1}^n X_i^r = E(X^r), \quad r = 1, \dots, k.$$

若参数本身就是期望或方差, 直接用

$$\bar{X} \approx E(X), \quad \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \approx \text{Var}(X)$$

往往更简洁. 注意矩估计方差时使用的是分母为 n 的二阶样本中心矩, 不是分母为 $n-1$ 的无偏样本方差.

注: [矩估计的统一步骤]

1. 数清未知参数个数, 确定需要几个独立方程;
2. 计算相应总体矩, 并写成参数的函数;
3. 令总体矩等于样本矩;
4. 解方程或方程组, 并确认结果只含样本和已知量.

使用前还要检查所选总体矩是否存在. 矩估计的合理性来自大数定律, 但不意味着它天然无偏或最有效.

7.2.2 一个参数的矩估计

例题 7.3 (连续型总体的矩估计与一致性)

总体密度为

$$f(x; \theta) = \begin{cases} \frac{\theta^2}{x^3} e^{-\theta/x}, & x > 0, \\ 0, & \text{其他}, \end{cases} \quad \theta > 0.$$

求 θ 的矩估计量, 并证明它具有一致性.

只有一个未知参数, 先计算一阶总体矩. 令 $u = \theta/x$, 可得

$$E(X) = \int_0^{\infty} x \frac{\theta^2}{x^3} e^{-\theta/x} dx = \theta.$$

令 $\bar{X} = E(X)$, 得到

$$\hat{\theta}_M = \bar{X}.$$

由于 $X_1, \dots, X_n$ 独立同分布且 $E(X_i) = \theta$ 存在, 辛钦大数定律给出

$$\hat{\theta}_M = \bar{X} \xrightarrow{P} \theta.$$

故 $\hat{\theta}_M$ 是 θ 的一致估计量.

方法选择: 一个参数优先尝试一阶矩方程. **结果检查:** θ 与 X 具有相同量纲, $\bar{X}$ 的量纲也相同.

例题 7.4 (从矩估计量到矩估计值)

设 $0 < \theta < 1/2$, 总体分布为

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

样本值为 3, 1, 3, 0, 3, 1, 2, 3, 求 θ 的矩估计值.

总体期望为

$$E(X) = 2\theta(1-\theta) + 2\theta^2 + 3(1-2\theta) = 3 - 4\theta.$$

令 $\bar{X} = 3 - 4\theta$, 得矩估计量

$$\hat{\theta}_M = \frac{3 - \bar{X}}{4}.$$

本组数据的样本均值为 $\bar{x} = 16/8 = 2$, 所以矩估计值为

$$\hat{\theta}_M = \frac{3 - 2}{4} = \frac{1}{4}.$$

结果检查: $1/4$ 落在参数空间 $(0, 1/2)$ 内.

7.2.3 两个参数的矩估计

例题 7.5 (正态总体两个未知参数)

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, μ, σ^2 均未知. 求二者的矩估计量.

可以令前两个原点矩相等:

$$\bar{X} = E(X) = \mu, \quad \frac{1}{n} \sum_{i=1}^n X_i^2 = E(X^2) = \sigma^2 + \mu^2.$$

解得

$$\hat{\mu}_M = \bar{X}, \quad \hat{\sigma}_M^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

也可直接用样本一阶原点矩估计期望、样本二阶中心矩估计方差, 结果相同.

易错提醒: 这里 $\hat{\sigma}_M^2$ 的分母为 n , 它一般不是无偏估计; 矩估计法追求的是矩方程匹配, 不自动保证无偏性.

小结

矩估计把样本矩与总体矩对应相等, 再解参数方程. 一个参数通常先用 $\bar{X} = E(X)$; 多个参数需要同样多个独立方程. 使用时先检查矩存在, 再检查解是否只含样本量, 并区分二阶样本中心矩与无偏样本方差.

7.3 最大似然估计法

7.3.1 核心思想: 让已观测数据尽可能“像是会出现的”

固定一组已经观测到的样本值 $x_1, \dots, x_n$. 不同参数会给这组数据不同的可能程度. 最大似然法选择使该组数据的似然函数最大的参数.

定义 7.6 (似然函数)

若总体为离散型, 概率质量函数为 $p(x; \theta)$, 则

$$L(\theta) = \prod_{i=1}^n p(x_i; \theta).$$

若总体为连续型, 密度为 $f(x; \theta)$, 则

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta).$$

$L(\theta)$ 称为样本观测值的**似然函数**.

连续型情形中, 单点概率为零, 所以 $L(\theta)$ 不是“恰好取这些点的概率”; 密度乘积反映的是这些观测点附近出现的相对可能程度. 似然函数中数据已固定, 变量是参数 θ .

定义 7.7 (最大似然估计)

若参数值 $\hat{\theta}$ 使

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta),$$

则称 $\hat{\theta}$ 为 θ 的**最大似然估计值**. 把小写样本值换回随机样本, 所得统计量称为**最大似然估计量**, 常记为 $\hat{\theta}_L$.

因对数函数严格递增, 最大化 $L(\theta)$ 与最大化对数似然

$$\ell(\theta) = \ln L(\theta)$$

等价. 取对数可把连乘变成求和, 便于求导.

注: [最大似然的标准流程]

1. 写清参数空间与样本必须满足的取值条件;
2. 构造似然函数 $L(\theta)$;
3. 取对数得到 $\ell(\theta)$;
4. 求驻点, 并结合参数空间、二阶导数或单调性判断最大值;
5. 把样本值 x_i 换成随机样本 X_i , 写出估计量.

若支持集依赖参数, 最大值可能出现在边界, 不能只会令导数为零.

7.3.2 连续型总体: 对数求导

例题 7.8 (连续型总体的最大似然估计)

仍设

$$f(x; \theta) = \frac{\theta^2}{x^3} e^{-\theta/x}, \quad x > 0, \text{quad} \theta > 0.$$

求 θ 的最大似然估计量.

对一组正的样本值 $x_1, \dots, x_n$,

$$L(\theta) = \prod_{i=1}^n \frac{\theta^2}{x_i^3} e^{-\theta/x_i} = \frac{\theta^{2n}}{\prod_{i=1}^n x_i^3} \exp\left(-\theta \sum_{i=1}^n \frac{1}{x_i}\right).$$

取对数:

$$\ell(\theta) = 2n \ln \theta - 3 \sum_{i=1}^n \ln x_i - \theta \sum_{i=1}^n \frac{1}{x_i}.$$

令一阶导数为零,

$$\ell'(\theta) = \frac{2n}{\theta} - \sum_{i=1}^n \frac{1}{x_i} = 0,$$

得到

$$\hat{\theta}_L = \frac{2n}{\sum_{i=1}^n 1/X_i}.$$

并且 $\ell''(\theta) = -2n/\theta^2 < 0$, 所以该驻点确为唯一最大值点.

关键转折: 矩估计计算 $E(X)$; 最大似然估计不积分, 而是把密度值连乘后对参数求极值.

7.3.3 离散型总体: 按取值频数归类

例题 7.9 (同一总体的矩估计与最大似然估计可以不同)

对前述分布

$$P(X = 0) = \theta^2, \text{quad} P(X = 1) = 2\theta(1 - \theta),$$

$$P(X = 2) = \theta^2, \text{quad} P(X = 3) = 1 - 2\theta, \text{quad} 0 < \theta < \frac{1}{2},$$

给定样本值 3, 1, 3, 0, 3, 1, 2, 3, 求最大似然估计值.

样本中 0, 1, 2, 3 的频数依次为 1, 2, 1, 4, 故

$$L(\theta) = (\theta^2)[2\theta(1-\theta)]^2(\theta^2)(1-2\theta)^4 = 4\theta^6(1-\theta)^2(1-2\theta)^4.$$

于是

$$\ell'(\theta) = \frac{6}{\theta} - \frac{2}{1-\theta} - \frac{8}{1-2\theta} = 0.$$

解得

$$\theta = \frac{7 \pm \sqrt{13}}{12}.$$

其中带加号的根不在 $(0, 1/2)$ 内, 故

$$\hat{\theta}_L = \frac{7 - \sqrt{13}}{12}.$$

结果检查: 必须用参数空间筛根. 本题矩估计值为 $1/4$, 与最大似然估计值不同, 这不表示其中一个计算错误, 而是两种原则不同.

7.3.4 分段密度与隐藏约束

例题 7.10 (先用密度规范性减少参数)

总体密度为

$$f(x; \alpha, \beta) = \begin{cases} \alpha, & -1 < x < 0, \\ \beta, & 0 < x < 1, \\ 0, & \text{其他}, \end{cases}$$

样本值为 $-0.5, 0.3, -0.2, -0.6, -0.1, 0.4, 0.5, -0.8$. 求 α 的最大似然估计值.

由密度规范性, $\alpha + \beta = 1$, 故 $\beta = 1 - \alpha$, 参数空间为 $0 < \alpha < 1$. 样本中有 5 个负值、3 个正值, 因此

$$L(\alpha) = \alpha^5(1-\alpha)^3.$$

令

$$\ell'(\alpha) = \frac{5}{\alpha} - \frac{3}{1-\alpha} = 0,$$

得

$$\hat{\alpha}_L = \frac{5}{8}.$$

方法选择: 看到多个参数, 先检查规范性等条件能否把它们联系起来; 分段密度则按样本落入各段的频数写似然.

7.3.5 支持集依赖参数: 最大值在边界

例题 7.11 (不能机械地令导数为零)

总体密度为

$$f(x; \theta) = \begin{cases} 2e^{-2(x-\theta)}, & x > \theta, \\ 0, & x \leq \theta. \end{cases}$$

求 θ 的最大似然估计量.

记 $x_{(1)} = \min(x_1, \dots, x_n)$. 只有当 $\theta < x_{(1)}$ 时似然为正, 并且

$$L(\theta) = 2^n \exp \left[-2 \sum_{i=1}^n x_i + 2n\theta \right].$$

在允许范围内, $L(\theta)$ 随 θ 严格递增, 所以最优参数应尽量靠近上边界 $x_{(1)}$. 按本课程的边界约定, 最大似然估计量写为

$$\hat{\theta}_L = X_{(1)} = \min(X_1, \dots, X_n).$$

严谨说明: 题面使用严格不等式 $x > \theta$ 时, $\theta = x_{(1)}$ 处似然为零, 严格地说只有当 $\theta \uparrow x_{(1)}$ 时取得上确界; 若支持集写成 $x \geq \theta$, 最大值就在边界取得. 考试按课程结论写 $X_{(1)}$, 同时理解这个边界细节.

7.3.6 两个未知参数与不变性

定理 7.12 (正态总体的最大似然估计)

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, 且 μ, σ^2 均未知, 则

$$\hat{\mu}_L = \bar{X}, \quad \hat{\sigma}_L^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

证明: 证明目标: 对正态似然的两个未知参数分别求偏导并解方程组.

出发点: 固定样本值后,

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right].$$

关键转换: 取对数, 把 σ^2 作为一个整体求偏导.

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

似然方程为

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0, \\ \frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0. \end{aligned}$$

先由第一式得 $\hat{\mu} = \bar{x}$, 再代入第二式得

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

换回随机样本即得结论.

证明小结: 两个参数对应两个偏导方程, 通常先解较简单的一个再回代. 方差的最大似然估计分母为 n , 与矩估计相同, 但不是无偏样本方差 S^2 . ■

定理 7.13 (最大似然估计的不变性)

若 $\hat{\theta}_L$ 是 θ 的最大似然估计, 且 g 为一函数, 则 $g(\hat{\theta}_L)$ 是 $g(\theta)$ 的最大似然估计.

例如, 若 $\hat{\theta}_L = \bar{X}$, 则 $3\bar{X} + 1$ 是 $3\theta + 1$ 的最大似然估计. 矩估计一般没有这样的不变性, 不能不经重新列矩方程就直接套用.

小结

最大似然估计固定观测数据, 把似然看成参数的函数并求最大值. 常规题用“写似然—取对数—求导—筛选参数空间—换回大写样本”; 分段密度要按频数归类, 支持集依赖参数时要检查单调性和边界, 不能只寻找驻点.

7.4 估计量的评价标准

得到估计量并不代表估计质量已经足够好. 本课程用无偏性、有效性和一致性从三个角度评价.

7.4.1 无偏性: 长期平均不系统偏高或偏低

定义 7.14 (无偏估计)

若对任意 $\theta \in \Theta$, 估计量 $\hat{\theta}$ 的期望存在且

$$E_{\theta}(\hat{\theta}) = \theta,$$

则称 $\hat{\theta}$ 是 θ 的无偏估计量.

无偏性描述重复抽样后估计量的平均位置, 不保证某一次估计一定接近真值, 也不保证方差小.

例题 7.15 (证明矩估计量无偏)

总体密度为

$$f(x; \theta) = \begin{cases} \frac{6x}{\theta^3}(\theta - x), & 0 < x < \theta, \\ 0, & \text{其他,} \end{cases} \quad \theta > 0.$$

求 θ 的矩估计量, 并判断其是否无偏.

计算总体期望:

$$E(X) = \int_0^{\theta} x \frac{6x}{\theta^3} (\theta - x) dx = \frac{\theta}{2}.$$

令 $\bar{X} = \theta/2$, 得

$$\hat{\theta} = 2\bar{X}.$$

于是

$$E(\hat{\theta}) = 2E(\bar{X}) = 2E(X) = \theta.$$

所以 $2\bar{X}$ 是 θ 的无偏估计量.

证明模式: 先明确待证等式 $E(\hat{\theta}) = \theta$, 再利用样本均值期望等于总体期望; 不要把“由矩估计得到”直接当成无偏的理由.

7.4.2 有效性: 同为无偏时比较波动

定义 7.16 (有效性比较)

设 $\hat{\theta}_1, \hat{\theta}_2$ 都是 θ 的无偏估计量, 方差均存在. 若

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2),$$

则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效.

例题 7.17 (三个无偏估计量的有效性比较)

设 X_1, X_2, X_3 是来自期望为 μ 、方差为 σ^2 的总体的简单随机样本. 比较

$$\hat{\mu}_1 = \bar{X}, \quad \hat{\mu}_2 = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3, \quad \hat{\mu}_3 = X_1$$

的有效性.

三组系数之和均为 1, 故三者期望均为 μ , 都是无偏估计. 利用样本独立性,

$$\begin{aligned} \text{Var}(\hat{\mu}_1) &= \frac{\sigma^2}{3}, \\ \text{Var}(\hat{\mu}_2) &= \left(\frac{1}{4} + \frac{1}{9} + \frac{1}{36}\right)\sigma^2 = \frac{7}{18}\sigma^2, \\ \text{Var}(\hat{\mu}_3) &= \sigma^2. \end{aligned}$$

因为

$$\frac{1}{3} < \frac{7}{18} < 1,$$

所以 $\hat{\mu}_1$ 最有效, $\hat{\mu}_3$ 最无效.

关键条件: 先确认无偏, 再比较方差; 线性组合的方差能写成系数平方和, 是因为样本相互独立.

7.4.3 一致性: 样本增多时逼近真值

定义 7.18 (一致估计量)

若对任意 $\theta \in \Theta$, 当 $n \rightarrow \infty$ 时

$$\hat{\theta}_n \xrightarrow{P} \theta,$$

则称 $\hat{\theta}_n$ 是 θ 的一致估计量, 也称相合估计量.

证明: 证明目标: 说明只要总体期望 $E(X) = \mu$ 存在, 样本均值就是 μ 的一致估计量.

出发点: 简单随机样本 $X_1, X_2, \dots$ 独立同分布, 且共同期望为 μ .

关键转换: 把 $\bar{X}_n$ 识别为独立同分布随机变量的算术平均.

由辛钦大数定律,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E(X) = \mu.$$

因此 $\bar{X}_n$ 是 μ 的一致估计量.

证明小结: 证明一致性不是计算某个固定 n 下的期望, 而是研究 $n \rightarrow \infty$ 的概率极限. 遇到样本均值或样本矩, 应优先检查大数定律条件. ■

标准	数学条件	回答的问题
无偏性	$E(\hat{\theta}) = \theta$	重复抽样的平均位置是否正确
有效性	无偏前提下方差更小	哪个无偏估计量波动更小
一致性	$\hat{\theta}_n \xrightarrow{P} \theta$	样本量增大后是否逼近真值

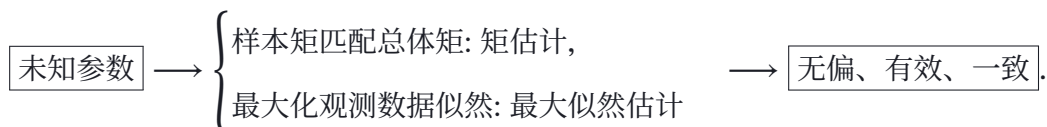
这三种标准不能互相替代: 无偏估计量可能方差很大; 有偏估计量也可能一致; 有效性比较在本章定义下以双方无偏为前提.

小结

无偏性看期望, 有效性在无偏前提下看方差, 一致性看样本量趋于无穷时的依概率收敛. 解题时先写对应定义式, 再选择期望运算、方差比较或大数定律, 不能只凭“看起来接近”下结论.

本章总结

一、知识结构



二、两种点估计方法对比

比较项	矩估计	最大似然估计
核心依据	样本矩逼近总体矩	已观测样本的似然应尽可能大
主要计算	求总体矩、列方程组	写似然、取对数、求极值
前提关注	所用总体矩必须存在	参数空间、支持集和独立性
多参数处理	建立同样多个矩方程	建立同样多个似然偏导方程
常见陷阱	把分母 n 写成 $n - 1$	忽略边界、参数范围或分段取值
性质	一般无不变性	一一变换下有不变性

三、方法选择与检查顺序

- 矩估计:** 数参数、选矩、算总体矩、列等式、解参数、查参数范围.
- 最大似然:** 定数据与参数角色、写支持条件、构造似然、取对数、求极值、筛参数空间、换回大写样本.
- 无偏性:** 直接计算 $E(\hat{\theta})$, 与 θ 比较.
- 有效性:** 先验证所有候选量无偏, 再比较方差.
- 一致性:** 写成依概率收敛目标, 优先寻找样本均值或样本矩并调用大数定律.

四、重点辨析与易错点

- 估计量是随机变量, 估计值是一次观测得到的数, 真实参数则固定但未知.
- 似然 $L(\theta; x_1, \dots, x_n)$ 与联合密度形式相同, 但用途不同: 数据固定, 参数变化.
- 最大似然的驻点只是候选点; 还要检查参数空间、二阶导数、端点和单调性.
- 连续型样本的似然是密度乘积, 不是若干单点概率的乘积.
- 正态方差的矩估计和最大似然估计均除以 n ; 无偏样本方差除以 $n - 1$.
- 无偏不等于误差小, 一致也不等于每个有限样本下都无偏.

五、证明掌握清单

- **A 级, 应能独立复现:** 根据矩估计思想建立方程并求解; 为离散型和连续型总体构造似然并处理驻点或单调边界; 用期望证明无偏; 在无偏前提下用方差比较有效性; 用大数定律证明一致性.
- **B 级, 应理解证明思路:** 样本矩为何能逼近总体矩; 最大似然为什么选择使已观察样本最可能出现的参数.
- **C 级, 知道结论即可:** 最大似然估计的一般相合性和渐近正态性; 充分性、完备性与克拉默-拉奥下界的证明. 这些内容本阶段不要求复现.

六、自测题

练习 7.19

用自己的话区分参数 θ 、估计量 $\hat{\theta}$ 和估计值, 并说明为什么评价无偏性时研究的是估计量.

练习 7.20

设总体 $X \sim P(\lambda)$, 求 λ 的矩估计量, 并证明它无偏且一致.

练习 7.21

设总体在区间 $(0, \theta)$ 上服从均匀分布. 分别求 θ 的矩估计量和最大似然估计量, 并说明最大似然为何来自样本最大值这一边界.

练习 7.22

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, μ, σ^2 均未知. 写出二者的矩估计和最大似然估计, 并判断 $\hat{\sigma}^2$ 是否无偏.

练习 7.23

若 X_1, X_2 独立同分布, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, 比较

$$\hat{\mu}_1 = \frac{X_1 + X_2}{2}, \quad \hat{\mu}_2 = \frac{3X_1 + X_2}{4}$$

的无偏性与有效性.

练习 7.24

某似然函数在参数空间内没有驻点, 但对参数严格递增. 说明下一步应检查什么, 为什么不能直接断言“最大似然估计不存在”.

完成自测后, 应能概括本章的统一逻辑: 先用样本构造参数估计量, 再分别从平均位置、波动大小和大样本极限评价它.

第八章 区间估计

本章导读

点估计只给参数一个估计值, 却没有直接说明误差范围和可靠程度. 区间估计在样本基础上构造一个随机区间, 并让这种构造方法以预先指定的比例覆盖参数真值.

本章的统一主线是

选枢轴量 $\longrightarrow$ 写中间概率 $\longrightarrow$ 代分位点 $\longrightarrow$ 解出参数区间.

看似繁多的正态总体公式, 都来自这四步. 首次学习应理解构造逻辑; 实际做题时还要熟练记忆公式、适用条件和自由度.

8.1 置信区间的概念

8.1.1 随机的是区间, 不是参数

设总体分布含有固定但未知的参数 θ , $X_1, \dots, X_n$ 为简单随机样本.

定义 8.1 (双侧置信区间)

若两个统计量

$$\hat{\theta}_L = \hat{\theta}_L(X_1, \dots, X_n), \quad \hat{\theta}_U = \hat{\theta}_U(X_1, \dots, X_n)$$

满足

$$P_{\theta}(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha,$$

则称随机区间 $(\hat{\theta}_L, \hat{\theta}_U)$ 为 θ 的置信水平为 $1 - \alpha$ 的双侧置信区间. 两端分别称为置信下限和置信上限.

$1 - \alpha$ 称为置信水平或置信度, 通常取 0.90, 0.95, 0.99 等较大的数; α 是未覆盖参数的总概率. 对连续型分布, 区间端点取开或闭不影响概率.

注: [95% 置信度的正确解释] 参数 θ 是固定的. 样本尚未抽取时, 区间端点是随机的; 反复独立抽样并用同一种方法构造许多区间, 其中约 95% 会覆盖真实参数. 得到某个具体区间后, 它要么覆盖真值, 要么不覆盖, 不能再说 “这个固定区间有 95% 的概率含有固定参数”.

例题 8.2 (理解重复抽样含义)

用置信水平 0.95 的方法独立构造 1000 个区间, 并不保证恰有 950 个覆盖真值; 覆盖个数是随机的, 但长期比例会接近 0.95. 置信水平评价的是**构造方法的长期覆盖率**, 不是某一个区间的优劣评分.

8.1.2 单侧置信限

若只关心参数的下界或上界, 可以构造单侧区间. 例如

$$P(\theta > \hat{\theta}_L) = 1 - \alpha$$

中的 $\hat{\theta}_L$ 是单侧置信下限. 本章与课程考查重点均为双侧置信区间; 单侧思想将在下一章单侧假设检验中再次出现.

小结

置信区间是由样本决定的随机区间, 置信水平描述重复抽样时的长期覆盖率. 解释具体结果时, 要牢记参数固定、区间随机; 使用双侧区间时还要确认 α 是否被正确分配到两侧.

8.2 区间估计的统一构造原理

8.2.1 枢轴量

定义 8.3 (枢轴量)

由样本和待估参数构成、其概率分布不含任何未知参数的随机变量, 称为枢轴量.

枢轴量与统计量不同: 统计量不能含未知参数, 而枢轴量通常**必须含待估参数**, 只是它的分布不能再含其他未知量. 区间估计正是先把待估参数放入一个分布已知的枢轴量, 再从概率不等

式中解出参数.

8.2.2 母型: 正态总体方差已知时估计均值

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, 其中 σ^2 已知、 μ 未知. 枢轴量为

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

本讲义统一使用上分位点: z_γ 满足

$$P(Z > z_\gamma) = \gamma.$$

标准正态分布对称, 因此

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

代入枢轴量并对不等式作等价变形, 得到

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

定理 8.4 (方差已知时正态均值的置信区间)

μ 的置信水平为 $1 - \alpha$ 的双侧置信区间为

$$\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right).$$

注: [为什么通常取等尾区间] 中间概率 $1 - \alpha$ 可以有多种分配方式. 本课程采用左右尾各为 $\alpha/2$ 的等尾区间; 在正态对称情形下, 它也给出以 $\bar{X}$ 为中心的最自然对称区间. 公式中的 $z_{\alpha/2}$ 是上 $\alpha/2$ 分位点, 不要误写成 z_α .

例题 8.5 (完整走一遍四步)

设总体 $X \sim N(\mu, 25)$, 抽取容量 $n = 100$ 的样本, 得到 $\bar{x} = 12.4$. 求 μ 的 95% 置信区间.

方差已知, 选枢轴量 $Z = (\bar{X} - \mu)/(5/\sqrt{100})$. 这里 $\alpha = 0.05$, 故 $z_{0.025} = 1.96$. 置信区间为

$$12.4 \pm 1.96 \times \frac{5}{10} = 12.4 \pm 0.98,$$

即

$$(11.42, 13.38).$$

结果检查: 区间以样本均值为中心; 标准差或置信水平增大时区间变宽, 样本量增大时区间变窄.

8.2.3 构造置信区间的五步模板

1. 明确总体类型、待估参数以及其他参数已知还是未知;
2. 选择包含待估参数且分布不含未知量的枢轴量;
3. 用上分位点写出中间概率为 $1 - \alpha$ 的不等式;
4. 对事件作等价变形, 把待估参数单独放在中间;
5. 代入样本观测值, 检查端点顺序、自由度和量纲.

后续公式均可按此模板推得. 按照本课程证明掌握要求, 应理解每个枢轴量为何可用和不等式怎样变形; 考试时不必从头重推全部公式.

小结

区间估计的核心不是背“加减哪一项”, 而是选择分布已知的枢轴量. 统一流程是定条件、选枢轴量、写中间概率、解参数、代数据; 本讲义所有分位点均采用上分位点记法.

8.3 一个正态总体的置信区间

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, 样本均值和样本方差分别为 $\bar{X}, S^2$.

8.3.1 均值的置信区间

当 σ^2 已知时, 使用标准正态枢轴量, 公式已经得到. 若 σ^2 未知, 不能把未知的 σ 留在端点中; 正态总体下应改用

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

定理 8.6 (方差未知时正态均值的置信区间)

当 σ^2 未知时, μ 的置信水平为 $1 - \alpha$ 的双侧置信区间为

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

例题 8.7 (未知方差要换成学生分布)

正态总体方差未知, 样本容量 $n = 16$, 得到 $\bar{x} = 20$, $s = 1$. 求 μ 的 90% 置信区间.

$\alpha = 0.10$, 自由度为 15, 故区间为

$$\left(20 - \frac{t_{0.05}(15)}{4}, 20 + \frac{t_{0.05}(15)}{4} \right).$$

若查得 $t_{0.05}(15)$ 的数值, 再代入即可.

方法选择: 未知方差时使用 S 和 $t(n-1)$, 不能把 S 代入正态公式后仍查 z 表.

8.3.2 方差的置信区间

估计 σ^2 时, 要先看 μ 是否已知.

若 μ 已知, 枢轴量为

$$Q = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n).$$

若 μ 未知, 需用样本均值代替, 枢轴量变为

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

卡方分布不对称. 采用上分位点时, 中间概率写成

$$P(\chi_{1-\alpha/2}^2(k) < Q < \chi_{\alpha/2}^2(k)) = 1 - \alpha.$$

注意 $\chi_{\alpha/2}^2(k)$ 是较大的右侧分位点.

定理 8.8 (正态方差的置信区间)

当 μ 已知时, σ^2 的置信水平为 $1 - \alpha$ 的双侧置信区间为

$$\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\alpha/2}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\alpha/2}^2(n)} \right).$$

当 μ 未知时, 区间为

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

注: [卡方区间为什么不是“中心加减”] 卡方分布只取正值且不对称, 区间端点来自除以两个不同的卡方分位点, 通常不以 S^2 为中心. 解不等式时因为 $\sigma^2 > 0$, 取倒数会改变不等号方向; 最稳妥的做法是写完后检查左端分母较大、右端分母较小.

例题 8.9 (先判断均值是否已知)

正态总体均值未知, 容量为 10 的样本得到 $s^2 = 4$. 则 σ^2 的 95% 置信区间为

$$\left(\frac{9 \times 4}{\chi_{0.025}^2(9)}, \frac{9 \times 4}{\chi_{0.975}^2(9)} \right).$$

结果检查: 未知均值消耗一个自由度, 所以分子是 $(n-1)s^2$, 两个卡方分位点的自由度均为 9.

8.3.3 一个总体的公式选择表

待估参数	已知条件	枢轴量	区间关键结构
μ	σ^2 已知	$(\bar{X} - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$	$\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$
μ	σ^2 未知	$(\bar{X} - \mu)/(S/\sqrt{n}) \sim t(n-1)$	$\bar{X} \pm t_{\alpha/2}(n-1)S/\sqrt{n}$
σ^2	μ 已知	$\sum (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$	平方和除以两个卡方分位点
σ^2	μ 未知	$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$	$(n-1)S^2$ 除以两个卡方分位点

小结

估计一个正态总体的均值时, 方差已知用 z , 方差未知用 t ; 估计方差时, 均值已知用自由度 n 的卡方平方和, 均值未知用自由度 $n - 1$ 的样本方差. 每次先判断“估谁、谁已知”, 再选公式.

8.4 两个正态总体的置信区间

设两组样本相互独立, 分别来自

$$N(\mu_1, \sigma_1^2), \quad N(\mu_2, \sigma_2^2),$$

样本容量为 n_1, n_2 , 样本均值和方差为 $\bar{X}, S_1^2$ 与 $\bar{Y}, S_2^2$.

8.4.1 均值差: 两个方差均已知

枢轴量为

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1).$$

定理 8.10 (已知方差时均值差的置信区间)

$\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的双侧置信区间为

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

这里“ $a \pm b$ ”表示区间 $(a - b, a + b)$.

8.4.2 均值差: 两个方差未知但相等

若 $\sigma_1^2 = \sigma_2^2$ 但共同方差未知, 定义合并样本方差

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

由第六章,

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2).$$

定理 8.11 (等方差未知时均值差的置信区间)

$\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的双侧置信区间为

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2}(n_1 + n_2 - 2)S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

注: 合并方差公式只能在两个正态总体方差相等的条件下使用. 本课程资料不要求处理“两个方差未知且不等”时的 Welch 区间; 若题目未说明等方差, 不能自行补上条件.

8.4.3 方差比

由第六章,

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

定理 8.12 (两个正态总体方差比的置信区间)

σ_1^2/σ_2^2 的置信水平为 $1 - \alpha$ 的双侧置信区间为

$$\left(\frac{S_1^2/S_2^2}{F_{\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{S_1^2/S_2^2}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)} \right).$$

例题 8.13 (按任务选公式)

两组独立正态样本容量分别为 $n_1 = 5, n_2 = 6$. 若比较总体方差, 样本方差为 s_1^2, s_2^2 , 则 95% 置信区间应写成

$$\left(\frac{s_1^2/s_2^2}{F_{0.025}(4, 5)}, \frac{s_1^2/s_2^2}{F_{0.975}(4, 5)} \right).$$

检查顺序: 分子样本对应第一个自由度 4, 分母样本对应第二个自由度 5; 交换方差比次序时, 自由度也要交换.

小结

两个总体问题先看待估的是均值差还是方差比. 均值差在方差已知时用 z , 方差未知但相等时用合并方差和 t ; 方差比统一用 F . 所有公式都要求两组样本相互独立, 且自由度顺序不能颠倒.

8.5 置信区间的精度与样本量

8.5.1 置信水平、区间长度和样本量的关系

对正态均值、方差已知的区间

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

半宽或允许误差为

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

区间总长度为 $2E$. 由此可见:

- 置信水平越高, α 越小, $z_{\alpha/2}$ 越大, 区间越宽;
- 总体标准差越大, 区间越宽;
- 样本量越大, 区间越窄, 但半宽只按 $1/\sqrt{n}$ 缩小.

可靠性和精度存在张力: 样本量不变时, 提高置信水平通常要牺牲区间长度; 若既要高置信度又要短区间, 就必须增加样本量.

8.5.2 由允许误差反推样本量

若要求半宽不超过给定值 $d > 0$, 则

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq d.$$

解得

$$n \geq \left(\frac{z_{\alpha/2} \sigma}{d} \right)^2.$$

样本量必须为整数, 因此最终应向上取整:

$$n_{\min} = \left\lceil \left(\frac{z_{\alpha/2} \sigma}{d} \right)^2 \right\rceil.$$

例题 8.14 (确定最小样本量)

某心理学家认为反应时间近似服从正态分布, 标准差可按 $\sigma = 0.05$ 秒处理. 希望以 95% 的置信水平估计平均反应时间, 并使误差不超过 0.01 秒. 至少需要多少个样本?

$z_{0.025} = 1.96$, 故

$$n \geq \left(\frac{1.96 \times 0.05}{0.01} \right)^2 = 96.04.$$

向上取整得到

$$n_{\min} = 97.$$

结果检查: 取 96 不满足 $n \geq 96.04$, 所以不能用通常的四舍五入; “区间长度不超过 0.02” 与 “半宽不超过 0.01” 是同一个要求.

小结

对已知方差的正态均值区间, 允许误差就是区间半宽. 反推样本量时先把 “总长度” 换成半宽, 再解不等式并向上取整; 置信度越高、波动越大或误差要求越严, 所需样本量越大.

本章总结

一、公式选择总表

待估参数	条件	置信区间中心或端点结构
单个 μ	σ^2 已知	$\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$
单个 μ	σ^2 未知	$\bar{X} \pm t_{\alpha/2}(n-1)S/\sqrt{n}$
单个 σ^2	μ 已知	$\sum(X_i - \mu)^2$ 分别除以 $\chi_{\alpha/2}^2(n)$ 、 $\chi_{1-\alpha/2}^2(n)$
单个 σ^2	μ 未知	$(n-1)S^2$ 分别除以 $\chi_{\alpha/2}^2(n-1)$ 、 $\chi_{1-\alpha/2}^2(n-1)$
$\mu_1 - \mu_2$	两方差已知	$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$
$\mu_1 - \mu_2$	两方差未知但相等	$\bar{X} - \bar{Y} \pm t_{\alpha/2}(n_1 + n_2 - 2)S_w\sqrt{1/n_1 + 1/n_2}$
σ_1^2/σ_2^2	两组独立正态样本	S_1^2/S_2^2 分别除以两个 F 上分位点

二、通用检查清单

1. 总体是否正态? 是一个总体还是两个独立总体?
2. 待估参数是均值、方差、均值差还是方差比?
3. 其他参数是已知、未知, 还是未知但相等?
4. 使用 z, t, χ^2, F 中哪一种枢轴量? 自由度是多少?
5. 本题给的是置信水平 $1 - \alpha$ 还是 α ? 双侧是否使用 $\alpha/2$?
6. 本讲义采用上分位点; 端点顺序是否正确?
7. 代入后区间是否满足量纲、正负和直观宽窄检查?

三、重点辨析与易错点

- 参数固定、区间随机; 具体区间不能解释为参数有某个概率落入其中.
- 枢轴量可以含待估参数, 但其分布不能含未知参数; 统计量则不能含任何未知参数.
- 方差未知时估计均值要用 t , 不能把 S 填入正态公式后仍查 z 表.
- 卡方和 F 分布不对称, 不能写成“中心加减同一个数”.
- 均值未知时估计方差, 自由度为 $n - 1$; 两个样本的 F 自由度有先后顺序.
- 反推样本量必须向上取整; 区间长度与允许误差相差 2 倍.

四、证明掌握清单

- **A 级, 应能准确写出并应用:** 课程范围内的全部置信区间公式; 根据总体、待估参数和已知条件选择分位点与自由度; 由半宽要求反推样本量并向上取整.
- **B 级, 应理解推导骨架:** “选枢轴量—写概率区间—代分位点—解参数区间”的统一流程; 三个母型及两个总体公式怎样由正态、 t 、卡方、 F 枢轴量得到; 置信水平、区间长度与样本量的关系; 重复抽样解释.
- **C 级, 知道结论即可:** 最短置信区间、枢轴量选择的一般理论、非正态总体的精确区间及渐近区间理论. 这些内容本阶段不要求复现.

五、自测题

练习 8.15

说明“某次算出的 95% 置信区间有 95% 概率包含固定参数”为什么不严谨, 并给出重复抽样解释.

练习 8.16

正态总体方差已知为 9, $n = 36$, $\bar{x} = 10$. 写出 μ 的 95% 置信区间, 并指出半宽.

练习 8.17

正态总体方差未知, $n = 25$, $\bar{x} = 8$, $s = 2$. 写出 μ 的 90% 置信区间, 说明所查分布和自由度.

练习 8.18

分别写出正态总体均值已知与未知时, σ^2 的 $1 - \alpha$ 置信区间, 并解释自由度为什么不同.

练习 8.19

两组独立正态样本方差未知但相等, 说明估计 $\mu_1 - \mu_2$ 时怎样构造 S_w^2 , 以及为何自由度为 $n_1 + n_2 - 2$.

练习 8.20

若已知 $\sigma = 4$, 希望以 99% 的置信度估计正态总体均值, 且误差不超过 1, 写出最小样本量的计算式并说明为什么向上取整.

完成自测后, 应能用一条主线统领全章: 先用已知抽样分布构造枢轴量, 再把中间概率事件等价变形形成参数的随机区间.

第九章 假设检验

本章导读

参数估计回答“参数大概是多少”，假设检验回答“样本是否提供了足够证据，反对关于总体的某个说法”。其基本逻辑是反证式的：先暂时假定原假设成立，再判断当前样本对应的结果是否过于反常。

全章可归结为一条流程：

提出 H_0, H_1 $\longrightarrow$ 选检验统计量 $\longrightarrow$ 由 H_1 定方向 $\longrightarrow$ 由 α 定临界值 $\longrightarrow$ 作结论。

第八章与本章使用相同的正态、 t 、卡方和 F 枢轴量；区别只是前者解出参数区间，本章则判断观测统计量是否落入拒绝域。

9.1 假设检验的基本思想

9.1.1 原假设与备择假设

定义 9.1 (统计假设)

关于总体分布形式或未知参数的命题称为统计假设。检验时成对写为

$$H_0 : \text{原假设}, \quad H_1 : \text{备择假设}.$$

H_0 是暂时保留、等待样本检验的命题； H_1 表示拒绝 H_0 后所支持的方向。

例如检验机器灌装均值是否仍为 500 克，可写

$$H_0 : \mu = 500, \quad H_1 : \mu \neq 500.$$

若只担心灌装不足，则备择假设应改为 $H_1 : \mu < 500$ 。备择方向改变，拒绝域方向也随之改变。

9.1.2 小概率事件原理

假定 H_0 成立后, 选择一个分布已知的检验统计量. 若观测值落入在 H_0 下概率很小的区域, 就认为 “ H_0 成立却出现如此极端数据” 不够可信, 从而拒绝 H_0 .

注: [小概率不等于不可能] 小概率事件在一次试验中仍可能发生. 因此假设检验不能给出绝对正确的逻辑证明, 而是在预先控制错误风险的前提下作统计决策. 拒绝 H_0 可能犯错; 未拒绝 H_0 也不表示 H_0 已被证明为真.

定义 9.2 (拒绝域与接受域)

在 H_0 成立时, 检验统计量取值中会导致拒绝 H_0 的区域称为**拒绝域**; 其余区域常称为接受域或非拒绝域. 分隔二者的点称为临界值.

更准确的结论措辞是:

- 统计量落入拒绝域: 在显著性水平 α 下**拒绝** H_0 ;
- 统计量未落入拒绝域: 在显著性水平 α 下**没有充分理由拒绝** H_0 .

9.1.3 两类错误与显著性水平

实际情况	检验决定	结果
H_0 为真	不拒绝 H_0	正确
H_0 为真	拒绝 H_0	第一类错误
H_0 为假	拒绝 H_0	正确
H_0 为假	不拒绝 H_0	第二类错误

定义 9.3 (两类错误)

第一类错误是 “弃真”, 其概率记为

$$\alpha = P(\text{拒绝}H_0 \mid H_0 \text{ 为真}).$$

第二类错误是 “取伪”, 其概率通常记为

$$\beta = P(\text{不拒绝}H_0 \mid H_0 \text{ 为假}).$$

给定的 α 称为**显著性水平**. 本课程的显著性检验优先控制第一类错误概率不超过 α . 样本量固定时, 不能通过简单移动临界值让 α 与所有备择参数下的 β 同时任意小; 增大样本量通常能改善两类风险.

9.1.4 标准检验步骤

1. 写出 H_0 与 H_1 , 明确是双侧、右侧还是左侧检验;
2. 在 H_0 成立条件下选择分布已知的检验统计量;
3. 根据显著性水平 α 和备择方向写拒绝域;
4. 代入样本观测值, 得到统计量的观测值;
5. 比较观测值与临界值, 并用完整句子作统计结论.

注: [备择假设决定拒绝域方向] 若 $H_1 : \theta \neq \theta_0$, 偏大或偏小都反对 H_0 , 所以拒绝域在两侧, 每侧概率为 $\alpha/2$; 若 $H_1 : \theta > \theta_0$, 只有统计量过大才反对 H_0 , 拒绝域在右侧; 若 $H_1 : \theta < \theta_0$, 拒绝域在左侧. 方向来自备择假设, 不来自样本观测值.

小结

假设检验先假定 H_0 成立, 再用小概率事件原理判断样本是否过于极端. H_1 决定拒绝域方向, α 控制第一类错误并决定临界值; 未落入拒绝域只能说证据不足, 不能证明 H_0 为真.

9.2 双侧、单侧检验与 p 值

9.2.1 三种拒绝域的统一形式

设在 H_0 下, 检验统计量 T 的分布关于 0 对称, 上 γ 分位点记为 t_γ . 则

备择假设	检验类型	拒绝域
$H_1 : \theta \neq \theta_0$	双侧	$ T > t_{\alpha/2}$
$H_1 : \theta > \theta_0$	右侧	$T > t_\alpha$
$H_1 : \theta < \theta_0$	左侧	$T < -t_\alpha$

对卡方、 F 等不对称且只取正值的分布, 左侧临界值应写成上 $1 - \alpha$ 分位点. 例如左侧 F 检验拒绝域为 $F < F_{1-\alpha}$.

例题 9.4 (备择方向不能由数据临时决定)

研究新工艺是否**提高**平均强度, 应事先写 $H_1: \mu > \mu_0$, 采用右侧检验. 即使样本均值恰好比 μ_0 小, 也不能看完数据后改成左侧检验; 这种事后改方向会破坏预先设定的第一类错误概率.

9.2.2 p 值

定义 9.5 (p 值)

在 H_0 成立的前提下, 得到“至少与当前观测结果同样极端”的统计量取值的概率, 称为该检验的 p 值. 等价地, p 值是刚好能够拒绝 H_0 的最小显著性水平.

决策规则为

$$p \leq \alpha \Rightarrow \text{拒绝 } H_0, \quad p > \alpha \Rightarrow \text{不拒绝 } H_0.$$

对标准正态检验, 若观测值为 z_0 , 则

$$H_1: \mu \neq \mu_0 \Rightarrow p = 2P(Z \geq |z_0|),$$

$$H_1: \mu > \mu_0 \Rightarrow p = P(Z \geq z_0),$$

$$H_1: \mu < \mu_0 \Rightarrow p = P(Z \leq z_0).$$

注: p 值不是“ H_0 为真的概率”, 也不是“本次结论犯错的概率”. 它是在 H_0 已被假定成立时, 当前数据及更极端数据出现的尾概率. p 值越小, 数据与 H_0 越不相容, 但效应大小还应结合实际背景判断.

9.2.3 与置信区间的联系

对使用同一枢轴量的双侧检验, 显著性水平为 α 的检验与置信水平为 $1 - \alpha$ 的置信区间具有对应关系:

$$\theta_0 \notin (1 - \alpha) \text{ 置信区间} \iff \text{拒绝 } H_0: \theta = \theta_0.$$

若 θ_0 落在区间内, 则不能拒绝 H_0 . 这种等价关系主要针对双侧检验, 且必须使用相同模型、枢轴量和显著性水平.

小结

双侧检验把 α 平分到两端, 右侧和左侧检验把全部 α 放在备择假设指向的一端. p 值用观测结果对应的尾概率度量反对 H_0 的证据强弱; 双侧检验还可用对应置信区间作等价判断.

9.3 一个正态总体均值的检验

设 $X_1, \dots, X_n$ 来自 $N(\mu, \sigma^2)$, 要检验

$$H_0 : \mu = \mu_0.$$

9.3.1 方差已知: Z 检验

在 H_0 下,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

定理 9.6 (正态均值的 Z 检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\mu \neq \mu_0$	$ Z > z_{\alpha/2}$
$\mu > \mu_0$	$Z > z_\alpha$
$\mu < \mu_0$	$Z < -z_\alpha$

例题 9.7 (完整的“统计量—临界值—结论”)

已知某正态总体标准差 $\sigma = 6$, 现取 $n = 36$, 得到 $\bar{x} = 102.5$. 在 $\alpha = 0.05$ 下检验

$$H_0 : \mu = 100, \quad H_1 : \mu > 100.$$

这是右侧 Z 检验, 统计量观测值为

$$z_0 = \frac{102.5 - 100}{6/\sqrt{36}} = 2.5.$$

右侧临界值 $z_{0.05} = 1.645$. 因为 $2.5 > 1.645$, 观测值落入拒绝域, 所以在 0.05 显著性水平下拒绝 H_0 , 认为样本提供了总体均值大于 100 的显著证据.

结果检查: 备择为“大于”, 拒绝域必须在右侧; 结论不能写成“证明 $\mu > 100$ ”.

9.3.2 方差未知: t 检验

若 σ^2 未知, 在 H_0 下使用

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1).$$

定理 9.8 (正态均值的 t 检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\mu \neq \mu_0$	$ T > t_{\alpha/2}(n-1)$
$\mu > \mu_0$	$T > t_{\alpha}(n-1)$
$\mu < \mu_0$	$T < -t_{\alpha}(n-1)$

注: [Z 与 t 怎样选] 选择标准只看总体方差是否已知, 不看样本量“大不大”. 本课程的精确结论建立在正态总体上: 方差已知用 σ 和 Z 检验; 方差未知用 S 和自由度为 $n-1$ 的 t 检验.

小结

一个正态总体均值检验以 $\bar{X} - \mu_0$ 为分子; 方差已知标准化为 Z, 方差未知学生化为 t. 备择假设为不等、大于、小于时, 分别使用双侧、右侧、左侧拒绝域.

9.4 一个正态总体方差的检验

要检验

$$H_0 : \sigma^2 = \sigma_0^2.$$

当 μ 已知时, 使用

$$Q = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \sim \chi^2(n).$$

当 μ 未知时, 使用

$$Q = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1).$$

下表用 k 表示相应自由度.

定理 9.9 (正态方差的卡方检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\sigma^2 \neq \sigma_0^2$	$Q < \chi_{1-\alpha/2}^2(k)$ 或 $Q > \chi_{\alpha/2}^2(k)$
$\sigma^2 > \sigma_0^2$	$Q > \chi_{\alpha}^2(k)$
$\sigma^2 < \sigma_0^2$	$Q < \chi_{1-\alpha}^2(k)$

其中 μ 已知时 $k = n$, μ 未知时 $k = n - 1$.

例题 9.10 (不对称分布的双侧拒绝域)

正态总体均值未知, $n = 20$, 要在显著性水平 0.05 下检验 $H_0 : \sigma^2 = 9$ 对 $H_1 : \sigma^2 \neq 9$.

若样本方差为 s^2 , 统计量为

$$q_0 = \frac{19s^2}{9}.$$

拒绝域是

$$q_0 < \chi_{0.975}^2(19) \quad \text{或} \quad q_0 > \chi_{0.025}^2(19).$$

易错提醒: 卡方分布不对称, 不能写成 $|Q| > c$; 左侧小临界值的上分位点下标是 0.975.

小结

正态方差检验使用卡方统计量. 均值已知时自由度为 n , 未知时为 $n - 1$; 双侧拒绝域位于两个不对称尾部, 右侧和左侧检验分别使用上 α 与上 $1 - \alpha$ 分位点.

9.5 两个正态总体均值差的检验

设两组独立样本分别来自 $N(\mu_1, \sigma_1^2)$ 与 $N(\mu_2, \sigma_2^2)$, 检验

$$H_0 : \mu_1 - \mu_2 = \delta_0.$$

9.5.1 两个方差已知

在 H_0 下,

$$Z = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1).$$

双侧、右侧、左侧拒绝域分别为

$$|Z| > z_{\alpha/2}, \quad Z > z_{\alpha}, \quad Z < -z_{\alpha}.$$

9.5.2 两个方差未知但相等

若 $\sigma_1^2 = \sigma_2^2$ 未知, 定义

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

则在 H_0 下

$$T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2).$$

双侧、右侧、左侧拒绝域分别为

$$|T| > t_{\alpha/2}(n_1 + n_2 - 2),$$

$$T > t_{\alpha}(n_1 + n_2 - 2), \quad T < -t_{\alpha}(n_1 + n_2 - 2).$$

例题 9.11 (检验两个工艺的平均值是否不同)

两组独立正态样本方差未知但可认为相等, 要检验

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2.$$

应取 $\delta_0 = 0$, 先由两组样本方差计算 S_w^2 , 再计算

$$t_0 = \frac{\bar{x} - \bar{y}}{s_w \sqrt{1/n_1 + 1/n_2}}.$$

若 $|t_0| > t_{\alpha/2}(n_1 + n_2 - 2)$, 则拒绝“两个均值相等”; 否则只能说尚无充分证据认为两个均值不同.

小结

两个均值差检验与其置信区间使用同一枢轴量. 两个方差已知用 Z ; 未知但相等用合并方差和自由度 $n_1 + n_2 - 2$ 的 t . 两组样本独立和等方差条件都不能漏掉.

9.6 两个正态总体方差比的检验

检验一般形式

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = r_0, \quad r_0 > 0.$$

在 H_0 下,

$$F = \frac{S_1^2/S_2^2}{r_0} \sim F(n_1 - 1, n_2 - 1).$$

定理 9.12 (两个正态总体方差比的 F 检验)

显著性水平为 α 时, 拒绝域为

H_1	拒绝域
$\sigma_1^2/\sigma_2^2 \neq r_0$	$F < F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$ 或 $F > F_{\alpha/2}(n_1 - 1, n_2 - 1)$
$\sigma_1^2/\sigma_2^2 > r_0$	$F > F_{\alpha}(n_1 - 1, n_2 - 1)$
$\sigma_1^2/\sigma_2^2 < r_0$	$F < F_{1-\alpha}(n_1 - 1, n_2 - 1)$

最常见的原假设是 $r_0 = 1$, 即 $H_0 : \sigma_1^2 = \sigma_2^2$, 此时检验统计量就是 S_1^2/S_2^2 .

注: [自由度顺序与比值顺序绑定] 分子 S_1^2 对应第一个自由度 $n_1 - 1$, 分母 S_2^2 对应第二个自由度 $n_2 - 1$. 若把统计量改写成倒数, 两个自由度和备择方向都要相应交换.

小结

方差比检验使用 F 分布. 原假设中的比值 r_0 要除入统计量; 双侧拒绝域在两个不对称尾部, 单侧方向仍由备择假设决定. 书写临界值时必须保持样本比值与自由度顺序一致.

9.7 统计结论与数学证明的边界

9.7.1 未拒绝不等于证明为真

一次样本未落入拒绝域, 可能因为 H_0 确实成立, 也可能因为样本量不足、数据波动较大而未能识别差异. 因此只能说“没有充分理由拒绝 H_0 ”, 不能说“接受并证明 H_0 为真”.

同理, 拒绝 H_0 表示数据在 H_0 下过于极端, 并承担至多 α 的第一类错误风险; 它不是演绎逻辑中的绝对否定.

9.7.2 反例与普遍命题

假设检验的证据逻辑不能替代数学命题的证明规则. 对全称命题“所有对象都具有性质 P ”:

- 找到一个不具有 P 的对象, 就能用反例否定全称命题;
- 检查若干对象都具有 P , 只能增加直观信心, 不能证明所有对象都具有 P ;
- 要证明普遍命题, 仍需从定义和已知条件作一般性推导.

例如, 要否定“任意两个不相关随机变量都独立”, 给出一个不相关但不独立的反例即可; 列举多个独立且不相关的例子, 却不能证明该命题成立.

小结

统计检验给出带错误风险的证据判断, 不是对参数命题的绝对证明. 数学上否定全称命题可用一个反例, 证明全称命题则必须覆盖一般对象; 这两种逻辑边界都应明确.

本章总结

一、统一决策流程

$H_0, H_1 \longrightarrow$ 在 H_0 下确定统计量分布 $\longrightarrow$ 按 H_1 定尾部 $\longrightarrow$ 按 α 定临界值 $\longrightarrow$ 比较并作结论.

二、正态总体检验选择表

待检参数	条件	检验统计量	原假设下分布
单个 μ	σ^2 已知	$(\bar{X} - \mu_0)/(\sigma/\sqrt{n})$	$N(0, 1)$
单个 μ	σ^2 未知	$(\bar{X} - \mu_0)/(S/\sqrt{n})$	$t(n-1)$
单个 σ^2	μ 已知	$\sum (X_i - \mu)^2/\sigma_0^2$	$\chi^2(n)$
单个 σ^2	μ 未知	$(n-1)S^2/\sigma_0^2$	$\chi^2(n-1)$
$\mu_1 - \mu_2$	两方差已知	标准化均值差	$N(0, 1)$
$\mu_1 - \mu_2$	两方差未知但相等	合并方差学生化均值差	$t(n_1 + n_2 - 2)$
σ_1^2/σ_2^2	两组独立正态样本	$(S_1^2/S_2^2)/r_0$	$F(n_1-1, n_2-1)$

三、方向与临界值速查

备择方向	对称分布	卡方或 F 分布	含义
$\neq$	$ T > c_{\alpha/2}$	小于上 $1 - \alpha/2$ 或大于上 $\alpha/2$ 分位点	双侧
$>$	$T > c_{\alpha}$	大于上 α 分位点	右侧
$<$	$T < -c_{\alpha}$	小于上 $1 - \alpha$ 分位点	左侧

四、重点辨析与易错点

- 备择假设决定拒绝域方向, 不能看完样本后再改成更容易拒绝的方向.
- 显著性水平是第一类错误概率的控制界, 不是 H_0 为假的概率.
- p 值是在 H_0 下的尾概率, 不是结论出错概率; $p \leq \alpha$ 才拒绝 H_0 .
- 未拒绝 H_0 不等于接受或证明 H_0 ; 结论必须保留“证据是否充分”的措辞.
- 卡方与 F 的双侧临界值不对称; 使用上分位点时, 下标顺序不能照搬对称分布.
- Z 、 t 、卡方、 F 的选择条件与上一章置信区间完全对应.

五、证明掌握清单

- **A 级, 应能独立完成:** 根据 H_0 选择检验统计量; 根据 H_1 推出双侧、左侧或右侧拒绝域; 说明 α 如何决定临界值; 完成“统计量—临界值—结论”的完整论证; 解释为何不能由未拒绝证明 H_0 ; 用反例否定全称命题.
- **B 级, 应理解思路:** 假设检验与置信区间使用相同枢轴量的联系; 两类错误与显著性水平; 单侧检验方向为何由备择假设决定.
- **C 级, 知道结论即可:** 各种拒绝域公式的重复代数推导; 检验最优性、功效函数与 Neyman–Pearson 理论. 这些内容本阶段不要求复现.

六、自测题

练习 9.13

分别说明第一类错误和第二类错误, 并解释为什么显著性水平 α 主要控制第一类错误.

练习 9.14

正态总体方差已知, 写出检验 $H_0: \mu = \mu_0$ 对三种备择 $\mu \neq \mu_0$ 、 $\mu > \mu_0$ 、 $\mu < \mu_0$ 的拒绝域.

练习 9.15

某检验得到 p 值 0.032. 分别在 $\alpha = 0.05$ 和 $\alpha = 0.01$ 下作结论, 并说明 p 值不能解释为什么.

练习 9.16

正态总体均值未知, 说明检验 $H_0: \sigma^2 = \sigma_0^2$ 时统计量、自由度和三类拒绝域.

练习 9.17

两组独立正态样本方差未知但相等, 写出检验 $H_0: \mu_1 - \mu_2 = 0$ 的统计量, 并说明双侧临界值与自由度.

练习 9.18

构造一个反例, 否定“任意两个不相关随机变量都相互独立”, 并说明为什么若干正例不能证明该全称命题.

完成自测后, 应能用一句话概括全章: 在预先控制第一类错误的前提下, 判断样本是否与原假设过于不相容, 并作出拒绝或不拒绝的证据结论.