

高等职业教育数学系列教材

概率统计学习教程

主 编	高文杰	张立圃
副主编	崔 媛	李国辉
主 审	陈秀岐	丁 杰

 天津大学出版社
TIANJIN UNIVERSITY PRESS

前 言

概率统计作为一门研究随机现象数量规律性的数学学科,它广泛应用于自然科学、技术科学、社会科学和人文科学的各个方面,掌握这门学科的基本理论和方法对提高数学素养和应用其知识认识与解决相关问题的能力是重要的.这门学科不仅具有理论的抽象性、逻辑的严谨性,而且还具有自己独特的思维方法和解题技巧,对初学者往往是原理能理解而习题做不出,表现为缺乏思路,不得要领.为帮助初学者学好这门学科,笔者根据多年从事概率统计教学所累积的经验,编写了此书.

全书内容在兼顾现行较通行范围的基础上,紧扣笔者编写的《概率统计》(天津大学出版社,丁杰、高文杰主编),故可作为其教学参考用书.同时还可作为高职高专及本科层次相关内容的学习辅导及教学参考.

全书共设9章,依次为概率论的基本概念、随机变量及其分布、随机变量的数字特征、随机向量、样本及抽样分布、参数估计、假设检验、方差分析和回归分析.每章以节为单元展开辅导,每节设有知识要点、典型题型解析、同步练习及参考答案,每章还设有疑难解析、常考题型归纳、知识网络与检测题和答案,书末还给出了数套期末测试题与答案,以供自我检测.

典型题型解析部分对常见的题型介绍了一般的解题思路方法,并精选例题给出详尽解答,力求体现典型例题的示范作用.疑难解析将概率统计中理解困难、易于混淆的重要概念、思想方法和解题方法深入剖析,有助于正确把握相关内容.常考题型归纳以表格形式对不同知识点的常考题型做出归类总结,有助于从整体上把握问题的不同侧面.每章的检测题和期末测试题都经过精心挑选,有许多是全国各类考试的优秀试题,这将有益于检验和提高解决问题的能力,巩固学习效果.

参加本书编写的有高文杰、张立圃、崔媛、李国辉,由高文杰统稿.陈秀岐、丁杰任主审.

给予本书很大帮助并提出宝贵意见的有杜俊文、李艳梅、陈洁、刘振云、王鲁静、李仲佳、周爱丽、蒋风光和徐向东等,编者在此一并致谢.

限于编者水平,本书一定存在疏漏和错误,恳请读者批评指正.

编者
2006年9月

目 录

第1章 概率论的基本概念	(1)
1.1 随机事件及其概率	(1)
1.2 事件的运算及概率的加法公式	(8)
1.3 条件概率与事件的独立性	(13)
1.4 全概率公式	(18)
疑难解析	(25)
本章常考题型归纳	(26)
本章知识结构网络图	(27)
检测题(一)	(27)
检测题(二)	(30)
第2章 随机变量及其分布	(33)
2.1 离散型随机变量	(33)
2.2 连续型随机变量	(39)
2.3 随机变量的分布函数与随机变量函数的分布	(45)
疑难解析	(60)
本章常考题型归纳	(62)
本章知识结构网络图	(63)
检测题(一)	(63)
检测题(二)	(67)
第3章 随机变量的数字特征	(70)
3.1 离散型随机变量的期望	(70)
3.2 连续型随机变量的期望	(72)
3.3 数学期望的性质与随机变量函数的期望公式	(74)
3.4 方差及其性质	(77)
疑难解析	(80)
本章常考题型归纳	(82)
本章知识结构网络图	(82)
检测题(一)	(82)
检测题(二)	(84)
第4章 随机向量	(87)
4.1 二维随机向量	(87)
4.2 二维随机向量的分布	(90)
4.3 二维随机向量函数的分布	(95)

4.4 二维随机向量的数字特征.....	(99)
4.5 大数定律与中心极限定理	(103)
疑难解析.....	(105)
本章常考题型归纳.....	(109)
本章知识结构网络图.....	(110)
检测题(一).....	(110)
检测题(二).....	(114)
第5章 样本及抽样分布.....	(117)
5.1 总体与样本	(117)
5.2 抽样分布	(118)
5.3 条形图、直方图及经验分布函数.....	(124)
疑难解析.....	(125)
本章常考题型归纳.....	(126)
本章知识结构网络图.....	(126)
检测题(一).....	(126)
检测题(二).....	(128)
第6章 参数估计.....	(130)
6.1 参数的点估计	(130)
6.2 估计量的评价标准	(136)
6.3 区间估计	(138)
本章常考题型归纳.....	(143)
本章知识结构网络.....	(143)
检测题(一).....	(143)
检测题(二).....	(145)
第7章 假设检验.....	(147)
7.1 假设检验	(147)
7.2 单个正态总体的假设检验	(150)
7.3 两个正态总体的假设检验	(157)
7.4 总体分布函数的假设检验	(166)
疑难解析.....	(172)
本章常考题型归纳.....	(175)
本章知识结构网络.....	(176)
检测题(一).....	(176)
检测题(二).....	(181)
第8章 方差分析.....	(185)
8.1 单因素方差分析	(185)
8.2 双因素方差分析	(191)
疑难解析.....	(197)
本章常考题型归纳.....	(198)

本章知识结构网络.....	(198)
检测题(一).....	(198)
检测题(二).....	(200)
第 9 章 回归分析.....	(202)
9.1 一元线性回归	(202)
9.2 多元线性回归	(211)
疑难解析.....	(216)
本章常考题型归纳.....	(218)
本章知识结构网络.....	(219)
检测题(一).....	(219)
检测题(二).....	(221)
期末测试及参考答案.....	(222)

第1章 概率论的基本概念

1.1 随机事件及其概率

1.1.1 知识要点

1. 随机事件

在一定条件下可能出现也可能不出现的现象称为随机事件,简称为事件,记为 $A, B, C, \dots$.

在一定条件下必然出现的现象称为必然事件,记为 Ω .

在一定条件下必然不出现的现象称为不可能事件,记为 $\emptyset$.

为了讨论问题方便起见,将必然事件 Ω 和不可能事件 $\emptyset$ 也看作随机事件.

2. 频率

①在一定的条件下,重复进行 n 次试验,在这 n 次试验中,随机事件 A 发生的次数 μ 称为随机事件 A 的频数,比值 $f_n(A) = \frac{\mu}{n}$ 称为随机事件 A 的频率.

大量的实践证明,随着试验次数 n 的逐渐增大,频率 $f_n(A)$ 逐渐稳定于一个常数 p ,这种性质称为频率的稳定性,常数 p 称为频率稳定值.频率的稳定性说明随机事件在试验中出现的可能性是可以度量的,即说明概率的存在性.

②基本性质:

$$0 \leq f_n(A) \leq 1.$$

3. 概率

①在一组不变的条件 S 下,随机事件 A 的频率稳定值 p 称为随机事件 A 在条件组 S 下发生的概率,记为 $P(A)$,即

$$P(A) = p.$$

②基本性质:

对任何随机事件 A 有 $0 \leq P(A) \leq 1$,

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

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

4. 排列与组合

(1)两个基本原理

加法原理:完成一件事情,有 m 类方法,第1类中有 n_1 种不同方法,第2类中有 n_2 种不

同方法 $\dots$ 第 m 类中有 n_m 种方法,任意一种方法都可完成这件事,那么完成这件事的不同方法共有 $n_1 + n_2 + \dots + n_m$ 种.

乘法原理:完成一件事情,有 m 个步骤,第一步有 n_1 种不同方法,第 2 步有 n_2 种不同方法 $\dots$ 第 m 步有 n_m 种不同方法,每一步都做完,这件事才完成,那么完成这件事的不同方法共有 $n_1 \cdot n_2 \cdot \dots \cdot n_m$ 种.

(2)排列

从 n 个不同元素中任取 r 个元素,按照一定顺序排成一列,称为从 n 个不同元素中取出 r 个元素的一个排列.

当 $r < n$ 时,称为选排列,排列总数用 A_n^r 来表示

$$A_n^r = n(n-1)\dots n(n-r+1) = \frac{n!}{(n-r)!};$$

当 $r = n$ 时,称为全排列,排列总数用 P_n 来表示

$$P_n = n(n-1)\dots 3 \cdot 2 \cdot 1 = n!$$

(3)组合

从 n 个不同元素中任取 r 个元素,不计顺序构成一组,称为从 n 个不同元素中任取 r 个元素的一个组合,其组合总数用 C_n^r 来表示,即

$$C_n^r = \frac{A_n^r}{r!} = \frac{n(n-1)\dots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}.$$

5. 古典概型

①如果一个事件组 $A_1, A_2, \dots, A_n$ 具有下列 3 条性质:

等可能性: $A_1, A_2, \dots, A_n$ 发生的机会相同;

完全性:在任一次试验中, $A_1, A_2, \dots, A_n$ 中至少有一个发生;

互不相容性:在任一次试验中, $A_1, A_2, \dots, A_n$ 中至多有一个发生.

则称事件组 $A_1, A_2, \dots, A_n$ 为等概基本事件组,或称为等可能完备事件组,其中任一事件 $A_i (i=1, 2, \dots, n)$ 称为基本事件.

②如果 $A_1, A_2, \dots, A_n$ 是一个等概基本事件组,而事件 B 由其中的某 m 个基本事件构成,则事件 B 的概率为

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

利用此公式来讨论事件的概率的模型称为古典概型.

1.1.2 典型题型解析

例 1.1.1 一次投掷两颗骰子,求出现的点数之和为奇数的概率.

解法 1 设随机事件 A = “出现的点数之和为奇数”.

若取每次投掷骰子所有可能出现的点数 (i, j) 为基本事件 $(i=1, 2, \dots, 6, j=1, 2, \dots, 6)$, 则基本事件总数 $n = C_6^1 \cdot C_6^1 = 36$, 并且 36 个基本事件组成等可能完备事件组,其中事件 A 包含的基本事件数 $m = 2C_3^1 \cdot C_3^1 + C_3^1 \cdot C_3^1 = 18$, 所以 $P(A) = \frac{18}{36} = \frac{1}{2}$.

解法2 可取每次投掷骰子可能出现的点数之和的奇偶性为基本事件,即基本事件为{点数之和为奇数},{点数之和为偶数},它们构成了一个等可能完备事件组,基本事件总数 $n=2$,事件 A 包含的基本事件数 $m=1$,所以 $P(A)=\frac{1}{2}$.

解法3 把每次投掷骰子可能出现的结果取为{奇,偶},{奇,奇},{偶,奇},{偶,偶},则这四个基本事件组成一个等可能完备事件组,基本事件总数 $n=4$,事件 A 包含的基本事件数 $m=2$,所以 $P(A)=\frac{2}{4}=\frac{1}{2}$.

注 本题说明同一问题可以选取不同的等可能完备事件组,但每个基本事件必须是等概的.

若{两个奇},{两个偶},{一奇一偶}作为基本事件,则会得出 $P(A)=\frac{1}{3}$,错误的原因就在于上述三个事件不是等概的,事实上, $P(\text{两个奇})=\frac{1}{4}\neq P(\text{一奇一偶})=\frac{1}{2}$.

例 1.1.2 k 个朋友随机地围绕圆桌而坐,求其中甲、乙座位相邻的概率.

解法1 设事件 A = “甲、乙两人座位相邻”.

k 个朋友随机地围绕圆桌而坐的所有可能的坐法共有 $k!$ 种,所以基本事件总数 $n=k!$,甲、乙两人座位要求相邻,可以分两步完成:先让甲、乙两人坐在一起,有 $2!$ 种坐法,其中 $(k-2)$ 个人随机地围绕圆桌而坐,共有 $(k-2)!$ 种坐法,由乘法原理,事件 A 包含的基本事件数 $m=2!(k-2)!$,所以 $P(A)=\frac{2!(k-2)!}{k!}=\frac{2}{k(k-1)}$.

解法2 设甲先坐好,再考虑乙的坐法,甲有 k 种坐法,乙有 $k-1$ 种坐法,由乘法原理,乙的坐法即基本事件总数 $n=k\cdot(k-1)$,而乙坐在甲身边的座法有 2 种,即事件 A 包含的基本事件数 $m=2$,所以 $P(A)=\frac{2}{k\cdot(k-1)}$.

例 1.1.3 袋中有 10 个球,9 个白球,1 个红球,10 个人依次从袋中各取一个球,每人取一球后不放回袋中,问第一人、第二人、... 最后一个人取到红球的概率各是多少?

解 设随机事件 A_k = “第 k 个人取得红球” ($k=1,2,\dots,10$),10 个人依次从袋中各取一球,每一种可能的取法对应于一个基本事件,基本事件总数 $n=10!$,事件 A_k 的发生意味着红球排在第 k 个位置,其余 9 个球排剩余的 9 个位置,即事件 A_k 包含的基本事件数为 $m=9!$,所以 $P(A)=\frac{9!}{10!}=\frac{1}{10}$. 答案与 k 无关,说明抽到红球的概率只与红球所占比率有关,而与第几次抽取无关.

例 1.1.4 共有 k 双不同型号的鞋子,从其中任意取出 $2r$ 只 ($2r\leq n$),求下列事件的概率:

- (1) A = “没有一双配对”;
- (2) B = “恰有一双配对”;
- (3) C = “恰有两双配对”;
- (4) D = “恰有 r 双配对”.

解 从 k 双鞋中取出 $2r$ 只的所有可能结果为 C_{2k}^{2r} 种,即基本事件总数 $n=C_{2k}^{2r}$.

(1)要取的 $2r$ 只鞋没有一双配对,只需选取 k 双中任意取出 $2r$ 双,再从每双中任取一只,共有 $C_k^{2r} \cdot (C_2^1)^{2r}$ 种取法,事件 A 包含的基本事件数 $m = C_k^{2r} \cdot (C_2^1)^{2r}$,所以

$$P(A) = \frac{C_k^{2r} \cdot (C_2^1)^{2r}}{C_{2k}^{2r}}.$$

(2)要取的 $2r$ 只鞋有 2 只配对,首先从 k 双中任取一双,有 C_k^1 种取法,再从剩下的 $k-1$ 双鞋中任取 $2r-2$ 双并从每双中取出一只,有 $C_{k-1}^{2r-2} \cdot (C_2^1)^{2r-2}$ 种取法,所以

$$P(B) = \frac{C_k^1 \cdot C_{k-1}^{2r-2} \cdot (C_2^1)^{2r-2}}{C_{2k}^{2r}}.$$

(3)本题与(2)题类似,所以 $P(C) = \frac{C_k^2 \cdot C_{k-2}^{2r-4} \cdot (C_2^1)^{2r-4}}{C_{2k}^{2r}}.$

(4)“恰有 r 双配对”意味着取出的 $2r$ 只鞋正好为 r 双,即从 k 双鞋中任取 r 双,有 C_k^r 种取法,所以 $P(D) = \frac{C_k^r}{C_{2k}^{2r}}.$

例 1.1.5 设每人生于一年中任意一个月都是等可能的,求下列事件的概率:

(1) $A =$ “12 个人的生日在 12 个不同的月份”;

(2) $B =$ “6 个人的生日恰在两个月中”.

解 (1)12 个人出生月份的所有可能有 12^{12} 种,即基本事件总数 $n = 12^{12}$. 12 个人的生日在 12 个不同的月份共有 $12!$ 种可能,即事件 A 包含的基本事件数 $m = 12!$,所以

$$P(A) = \frac{12!}{12^{12}} \approx 0.000\ 054.$$

(2)6 个人出生月份的所有可能有 12^6 种,即基本事件总数 $n = 12^6$.

要使 6 个人的生日恰在两个月中,必须先从 12 个月中任选两个月,共有 C_{12}^2 种选法;然后 6 个人的生日月份在这两个月中选择,但必须减去都在某一个月中的情况,即有 $2^6 - 2$ 种,因此事件 B 包含的基本事件数 $m = C_{12}^2 \cdot (2^6 - 2)$,所以 $P(B) = \frac{C_{12}^2 \cdot (2^6 - 2)}{12^6} \approx 0.001\ 37$.

例 1.1.6 从 1 2 ... 10 共 10 个数中任取一数,设每个数以 $\frac{1}{10}$ 的概率被取到,取后放回,共取 7 个数,求下列事件的概率:

(1) $A =$ “7 个数全不相同”;

(2) $B =$ “不含 10 与 1”;

(3) $C =$ “10 恰好出现两次”;

(4) $D =$ “取到的最大数恰好为 6”;

(5) $E =$ “取出的 7 个数形成一个严格上升序列”.

解 根据题意可知,基本事件总数 $n = 10^7$.

(1)取出的 7 个数都不相同,根据排列的定义可知,事件 A 包含的基本事件数 $m = A_{10}^7$,所以

$$P(A) = \frac{A_{10}^7}{10^7} \approx 0.060\ 5.$$

(2)因为取出的 7 个数不含 10 与 1,所以在剩下的 8 个数中有放回地取 7 个数,因此事件 B 包含的基本事件数 $m = 8^7$,所以 $P(B) = \frac{8^7}{10^7} \approx 0.209\ 7$.

(3) 10 出现两次可以是 7 次中的任意两次, 有 C_7^2 种选法, 在其他 5 次中, 每次可以取其 9 个数中的任意一个, 有 9^5 种选法, 因此事件 C 包含的基本事件数 $m = C_7^2 \cdot 9^5$, 所以

$$P(C) = \frac{C_7^2 \cdot 9^5}{10^7} \approx 0.124.$$

(4) 从 1, 2, ..., 10 这 10 个数中取 7 个数, 其中最大数不大于 6 的取法有 6^7 , 最大数不大于 5 的取法有 5^7 种, 因此, 事件 D 包含的基本事件数 $m = 6^7 - 5^7$, 所以 $P(D) = \frac{6^7 - 5^7}{10^7} \approx 0.0202$.

(5) 若取出的 7 个数形成一个严格上升序列, 则 7 个数必然完全不相同, 7 个不同号码有 $7!$ 种不同排列, 其中只有一个是按严格上升次序的排列, 也就是说, 一种组合对应一种严格上升排列, 因此事件 E 包含的基本事件数 $m = C_{10}^7$, 所以 $P(E) = \frac{C_{10}^7}{10^7} \approx 0.000012$.

例 1.1.7 电梯从第 1 层到第 14 层, 一开始电梯里面有 10 名乘客, 每位乘客都可能在 2 ~ 14 楼下电梯, 求下列事件的概率:

- (1) $A =$ “10 人在同一层下”;
- (2) $B =$ “10 人不在同一层下”;
- (3) $C =$ “10 人都在第 2 层下”;
- (4) $D =$ “10 人中有 4 人在第 10 层下”.

解 10 名乘客中的任意一名乘客都等可能在 2 ~ 14 层的 13 层楼中任选一楼下电梯, 因此基本事件总数 $n = 13^{10}$.

(1) 10 个人在同一层楼下电梯有 13 种可能, 即事件 A 包含的基本事件数 $m = 13$, 所以

$$P(A) = \frac{13}{13^{10}} = 13^{-9}.$$

(2) 10 个人不在同一层下, 意味着第 1 人在 13 层中任选一层下, 第二个人在余下的 12 层中任选一层下, ..., 第 10 人在余下的 4 层中任选一层下, 因此事件 B 包含的基本事件 $m = 13 \times 12 \times 11 \times \dots \times 4$, 所以 $P(B) = \frac{13 \times 12 \times \dots \times 4}{13^{10}}$.

(3) 作法类似于 (1) 题, 所以 $P(C) = \frac{1}{13^{10}}$.

(4) 10 个人中有 4 人在第 10 层下, 意味着先从 10 个人中任选 4 个人在第 10 层下, 有 C_{10}^4 种方法, 余下 6 人等可能地在余下的 12 层中任选一层下, 有 12^6 种方法, 因此事件 D 包含的基本事件数 $m = 2C_{10}^4 \cdot 12^6$, 所以 $P(D) = \frac{C_{10}^4 \cdot 12^6}{13^{10}}$.

例 1.1.8 20 名运动员中有 2 名种子选手, 现将运动员平均分成两组, 求下列事件的概率:

- (1) $A =$ “两名种子选手分在不同组”;
- (2) $B =$ “两名种子选手分在同一组”.

解 将 20 名运动员平均分成两组, 基本事件总数 $n = C_{20}^{10}$.

(1) 事件 A 包含的基本事件数 $m = C_2^1 \cdot C_{18}^9 \cdot C_{10}^{10} = 2C_{18}^9$, 其中 C_2^1 表示两名种子选手中有一人在此组, 剩下的 9 名选手从 18 名非种子选手中选, 为 C_{18}^9 , 剩下 10 名选手组成另一组, 为

C_{10}^{10} . 所以 $P(A) = \frac{C_2^1 \cdot C_{18}^9 \cdot C_{10}^{10}}{C_{20}^{10}} \approx 0.526$.

(2) 两名种子选手在同一组, 则该组在余下 18 名选手中选 8 名进行组合, 但因为有两组, 因此事件 B 包含的基本事件数 $m = 2C_{18}^8$, 所以 $P(B) = \frac{2C_{18}^8}{C_{20}^{10}} \approx 0.474$.

例 1.1.9 袋中有 1, 2, ..., N 号球各一个, 采用 (1) 无放回; (2) 有放回两种方式摸球, 求在第 k 次摸球时首次摸到 1 号球的概率.

解 设事件 A_k = “第 k 次摸球时首次摸到 1 号球”.

(1) 无放回摸球时. 从 N 个球中任取 k 个球的每个排列就是一个基本事件, 则基本事件总数 $n = A_N^k$. 事件 A_k 可分成两步完成: 首先在第 k 个位置上排上一号球, 有一种排法, 然后在前 $k-1$ 个位置上排上其他 $N-1$ 个球, 有 A_{N-1}^{k-1} 种排法, 由乘法原理事件 A_k 包含的基本事件数 $m = 1 \times A_{N-1}^{k-1}$, 所以 $P(A_k) = \frac{A_{N-1}^{k-1}}{A_N^k} = \frac{1}{N}$.

(2) 有放回摸球时. 基本事件总数 $n = N^k$, 事件 A_k 可分为两步完成: 首先前 $k-1$ 次未摸到 1 号球, 有 $(N-1)^{k-1}$ 种可能, 然后第 k 次摸到 1 号球有一种可能, 由乘法原理事件 A_k 包含的基本事件数 $m = (N-1)^{k-1} \times 1$, 所以 $P(A_k) = \frac{(N-1)^{k-1}}{N^k}$.

例 1.1.10 从一副扑克牌中(共 52 张)一张一张地取牌, 求下列事件的概率:

(1) A = “第 r 次取牌时, 第 1 次取出 A 牌”;

(2) B = “第 r 次取牌时, 第 2 次取出 A 牌”;

(3) C = “第 r 次取牌时, 第 3 次取出 A 牌”;

(4) D = “第 r 次取牌时, 第 4 次取出 A 牌”.

解 从 52 张牌中无放回地取出 r 张牌的所有可能结果有 A_{52}^r 种, 即基本事件总数 $n = A_{52}^r$.

(1) 事件 A 表明前 $r-1$ 次取牌均未取出 A 牌, 有 A_{48}^{r-1} 种可能, 第 r 次取出 A 牌, 有 C_4^1 种可能, 因此, 事件 A 包含的基本事件数 $m = A_{48}^{r-1} \cdot C_4^1$, 所以 $P(A) = \frac{A_{48}^{r-1} \cdot C_4^1}{A_{52}^r} = \frac{4A_{48}^{r-1}}{A_{52}^r}$.

(2) 事件 B 表明前 $r-1$ 次取牌恰有 1 次取出 A 牌, 有 $C_4^1 \cdot C_{r-1}^1 \cdot A_{48}^{r-2}$ 种可能, 第 r 次取牌时第 2 次取出 A 牌, 有 C_3^1 种可能, 因此, 事件 B 包含的基本事件数 $m = C_4^1 \cdot C_{r-1}^1 \cdot A_{48}^{r-2} \cdot C_3^1$, 所以

$$P(B) = \frac{C_4^1 \cdot C_{r-1}^1 \cdot A_{48}^{r-2} \cdot C_3^1}{A_{52}^r} = \frac{12(r-1)A_{48}^{r-2}}{A_{52}^r}.$$

(3) 事件 C 表明前 $r-1$ 次取牌中取出 2 张 A 牌与 $r-3$ 张非 A 牌, 再将这 $r-1$ 张牌进行排列, 共有 $C_4^2 \cdot C_{48}^{r-3} \cdot P_{r-1}$ 种排法, 第 r 次取牌时第 3 次取出 A 牌, 有 C_2^1 种可能, 因此事件 C 包含的基本事件数 $m = C_4^2 \cdot C_{48}^{r-3} \cdot P_{r-1} \cdot C_2^1$, 所以

$$P(C) = \frac{C_4^2 \cdot C_{48}^{r-3} \cdot P_{r-1} \cdot C_2^1}{A_{52}^r} = \frac{12(r-1)! A_{48}^{r-3}}{A_{52}^r}.$$

$$(4) \text{ 本题与(3)类似, 所以 } P(D) = \frac{4(r-1)! C_{48}^{r-4}}{A_{52}^r}.$$

同步练习 1.1

1. 10 件产品中有次品 3 件, 从这 10 件产品中一件一件地无放回地任意取出 4 件, 求 4 件中恰有 1 件次品的概率.

2. 从 5 双不同颜色的袜子中取 4 只, 求 4 只袜子中至少有 2 只配成一双的概率.

3. 任意将 10 本书放在书架上, 其中有两套书, 一套 3 卷, 一套 4 卷, 求下列事件的概率.

$A =$ “3 卷一套的放在一起”;

$B =$ “两套各自放在一起”.

4. 袋内放有两个 5 分的钱币, 三个 2 分的钱币, 五个 1 分的钱币, 任取 5 个, 求总数超过一角的概率.

5. 10 个人分别佩戴从 1 号到 10 号的胸卡, 任选 3 人记录其胸卡的号码, 求最小号码为 5 的概率.

6. 掷三颗骰子, 求所得点数能构成等差数列的概率.

7. 从 0~9 共 10 个数中每次任取一个数, 求:

(1) 无放回地取两个数, 恰为相邻数的概率.

(2) 有放回地取两个数, 恰为相邻数的概率.

8. 从一副扑克牌(52 张)中任取 5 张, 求下列事件的概率:

$A =$ “5 张牌恰为同一花色”;

$B =$ “3 张牌点数相同, 另两张是相同的另一点数”;

$C =$ “5 张牌中有两个不同的对”.

9. 袋中有编号为 1~5 的 5 个球, 从袋中有放回地取 3 个球, 求下列事件的概率:

$A =$ “取出的 3 个球的次序是 1、2、3”;

$B =$ “3 个球恰好是 1、2、3”;

$C =$ “3 个球恰好不同”.

10. 田径场上有 7 条跑道, 其中有 3 条好跑道, 7 名运动员抽签决定自己的跑道, 运动员甲最先抽, 乙随后抽. 问甲、乙两名运动员抽到好跑道的概率是否相等? 并证明你的结论.

同步练习 1.1 参考答案

$$\begin{aligned} 1. \frac{C_4^1 \cdot A_7^3 \cdot A_3^1}{A_{10}^4} &= \frac{1}{2}. 2. \frac{C_5^1 \cdot C_4^1 \cdot C_2^1 \cdot C_2^1 + C_5^2}{C_{10}^4} = \frac{13}{21}. 3. P(A) = \frac{8! \cdot 3!}{10!} = \frac{1}{15}, P(B) = \frac{5! \cdot 4! \cdot 3!}{10!} \\ &= \frac{1}{210}. 4. \frac{126}{C_{10}^5} = \frac{1}{2}. 5. \frac{10}{120} = \frac{1}{12}. 6. \frac{6 \times 3!}{6^3} = \frac{1}{6}. 7. \frac{9C_2^1}{C_{10}^1 \cdot C_9^1} = \frac{1}{5}, \frac{9C_2^1}{C_{10}^1 \cdot C_{10}^1} = \frac{9}{50}. 8. P(A) = \frac{C_4^1 \cdot C_{52}^5}{C_{52}^5} \\ &= \frac{33}{166\ 600}, P(B) = \frac{C_{13}^1 \cdot C_4^3 \cdot C_{12}^1 \cdot C_4^2}{C_{52}^5} = \frac{6}{4\ 165}, P(C) = \frac{C_{13}^2 \cdot C_4^2 \cdot C_4^2 \cdot C_{44}^1}{C_{52}^5} = \frac{198}{4\ 165}. 9. P(A) = \frac{1}{5^3}, \\ P(B) &= \frac{6}{5^3}, P(C) = \frac{A_5^3}{5^3} = \frac{60}{5^3}. 10. \text{相等. 因为甲、乙两名运动员抽到好跑道的概率都是 } \frac{3}{7}. \end{aligned}$$

1.2 事件的运算及概率的加法公式

1.2.1 知识要点

1. 事件间的关系

关系	符号	概率论	集合论
包含关系	$A \subset B$	如果事件 A 发生了, 则事件 B 必发生	集合 A 是集合 B 的子集
相等关系	$A = B$	事件 A 与事件 B 相等	集合 A 与集合 B 相等
对立关系	$\bar{A}$	事件 A 的对立事件	集合 A 的余集
互不相容	$AB = \emptyset$	事件 A 与事件 B 不能同时发生	集合 A 与集合 B 没有公共元素

2. 事件间的运算

运算	符号	概率论	集合论
和	$A + B$ (或 $A \cup B$)	事件 A 与事件 B 至少有一个发生	集合 A 与集合 B 的并集
积	AB	事件 A 与事件 B 同时发生	集合 A 与集合 B 的交集
差	$A - B$	事件 A 发生而事件 B 不发生	集合 A 与集合 B 的差集

3. 事件的运算律

交换律 $A + B = B + A$; $AB = BA$.

结合律 $(A + B) + C = A + (B + C)$; $(AB)C = A(BC)$.

分配律 $A + BC = (A + B)(A + C)$; $A(B + C) = AB + AC$.

对偶律 $\overline{A + B} = \bar{A}\bar{B}$; $\overline{AB} = \bar{A} + \bar{B}$.

4. 事件的性质

$A + A = A$, $A + \bar{A} = \Omega$, $A + \emptyset = A$, $A + \Omega = \Omega$,

$AA = A$, $A\emptyset = \emptyset$, $A\bar{A} = \emptyset$, $A\Omega = A$,

$\overline{(\bar{A})} = A$, $A - B = A\bar{B}$, $\bar{A} = \Omega - A$.

5. 概率的加法公式

(1) 若 $AB = \emptyset$, 则 $P(A + B) = P(A) + P(B)$,

特别地, $P(A) = 1 - P(\bar{A})$.

(2) 若 $A_1, A_2, \dots, A_n$ 是两两互不相容的事件, 则

$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$ (概率的有限可加性).

(3) 若 A, B 是两个任意事件, 则 $P(A + B) = P(A) + P(B) - P(AB)$.

(4) 若 A, B, C 是三个任意事件, 则

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

(5) 若 $A_1, A_2, \dots, A_n$ 是一事件列, 且两两互不相容, 则

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \text{ (概率的完全可加性).}$$

1.2.2 典型题型解析

1. 随机事件间的关系及其运算

例 1.2.1 设袋内有 10 个编号为 1~10 的球,从中任取一个,观察其号码,设 A = “取得球的号码是奇数”, B = “取得球的号码是偶数”, C = “取得球的号码小于 5”, 则

(1) $A+B$, AB , $\bar{C}$, $\overline{AC}$, $\overline{B+C}$, $\overline{BC}$, $A-C$ 各表示什么事件?

(2) 事件 AC 与 $\overline{AC}$ 是否不相容? 是否对立?

解 设 H_i = “取得球的号码为 i ” ($i=1, 2, \dots, 10$).

(1) $A+B$ = “取得球的号码或是奇数或是偶数”, 这是必然事件, 即 $A+B=\Omega$.

AB = “取得球的号码既是奇数又是偶数”, 这是不可能事件, 即 $AB=\emptyset$.

$\bar{C}$ = “取得球的号码大于或等于 5”, 即 $C = H_1 + H_2 + H_3 + H_4$, $\bar{C} = H_5 + H_6 + H_7 + H_8 + H_9 + H_{10}$.

$\overline{AC}$ = “取得球的号码是大于 5 的偶数”, 即 $\overline{AC} = H_6 + H_8 + H_{10}$.

$\overline{B+C} = \overline{BC}$ = “取得球的号码是大于或等于 5 的奇数”, 即

$$\overline{B+C} = H_5 + H_7 + H_9.$$

$\overline{BC} = \bar{B} + \bar{C}$ = “取得球的号码是奇数或者大于等于 5”, 即 $\overline{BC} = H_1 + H_3 + H_5 + H_6 + H_7 + H_8 + H_9 + H_{10}$.

$A-C = \overline{AC}$ = “取得球的号码是大于等于 5 的奇数”, 即 $A-C = H_5 + H_7 + H_9$.

(2) AC = “取得的球的号码是小于 5 的奇数”, 即 $AC = H_1 + H_3$, 而 $\overline{AC} = H_2 + H_4 + H_6 + H_8 + H_{10}$, $(AC)(\overline{AC}) = \emptyset$, $AC + \overline{AC} \neq \Omega$, 所以事件 AC 与事件 $\overline{AC}$ 互不相容, 但不对立.

例 1.2.2 设 A, B, C 为三个事件, 用 A, B, C 表示下列事件:

H_1 = “三个事件中恰有一个发生”;

H_2 = “三个事件中恰有两个发生”;

H_3 = “三个事件至少有一个发生”;

H_4 = “三个事件都不发生”;

H_5 = “三个事件不多于一个发生”;

H_6 = “三个事件不多于两个事件发生”.

解 $H_1 = \overline{ABC} + \overline{A}BC + A\overline{BC}$;

$$H_2 = \overline{ABC} + \overline{A}BC + A\overline{BC};$$

$$H_3 = A + B + C$$

$$= \overline{ABC} + \overline{A}BC + \overline{A}\overline{B}C + A\overline{BC} + A\overline{B}\overline{C} + \overline{A}BC + ABC;$$

$$H_4 = \overline{ABC}$$

$$= \overline{A+B+C};$$

$$H_5 = \overline{ABC} + \overline{A}BC + \overline{A}\overline{B}C + \overline{A}\overline{B}\overline{C}$$

$$= \overline{ABC} + \overline{A}BC + \overline{A}\overline{B}C + \overline{A}\overline{B}\overline{C}$$

$$= \overline{AB+AC+BC};$$

$$H_6 = \overline{ABC} + \overline{A}BC + \overline{A}\overline{B}C + \overline{A}\overline{B}\overline{C} + A\overline{BC} + A\overline{B}\overline{C} + \overline{A}BC$$

$$\begin{aligned}
 &= \overline{ABC} \\
 &= \overline{A} + \overline{B} + \overline{C}.
 \end{aligned}$$

例 1.2.3 化简下列各式：

$$(1)(A+B)(A+\overline{B});$$

$$(2)(A+B)(B+C);$$

$$(3)(A+B)(A+\overline{B})(\overline{A}+B).$$

$$\begin{aligned}
 \text{解 } (1)(A+B)(A+\overline{B}) &= (A+B)A + (A+B)\overline{B} = A + BA + A\overline{B} \\
 &= A + (BA + \overline{B}A) = A + A(B + \overline{B}) = A + A\Omega = A + A = A;
 \end{aligned}$$

$$\begin{aligned}
 (2)(A+B)(B+C) &= (A+B)B + (A+B)C = AB + B + AC + BC \\
 &= (AB + BC) + B + AC = ((AB + BC) + B) + AC = B + AC;
 \end{aligned}$$

$$\text{注 } AB + BC \subset B;$$

$$(3)(A+B)(A+\overline{B})(\overline{A}+B) = A(\overline{A}+B) = A\overline{A} + AB = AB.$$

注 此题利用了(1)的结果.

例 1.2.4 设 A, B 是两个事件, 证明下列等式：

$$(1)A+B=A+B\overline{A};$$

$$(2)\overline{A}B + A\overline{B} + \overline{A}\overline{B} = \overline{A}\overline{B};$$

$$(3)(A-B)(B-A) = \overline{AB} + \overline{A\overline{B}};$$

$$(4)P(A-B) = P(A) - P(AB).$$

$$\text{证明 } (1)A+B = (A+B)\Omega = (A+B)(A+\overline{A}) = A+B\overline{A}.$$

$$\begin{aligned}
 (2)\overline{A}B + A\overline{B} + \overline{A}\overline{B} &= (\overline{A}B + \overline{A}\overline{B}) + A\overline{B} = \overline{A}(B+\overline{B}) + A\overline{B} \\
 &= \overline{A}\Omega + A\overline{B} = \overline{A} + A\overline{B} = (\overline{A}+A)(\overline{A}+\overline{B}) = \overline{A}+\overline{B} = \overline{A\overline{B}}.
 \end{aligned}$$

$$\begin{aligned}
 (3)\overline{AB} + \overline{A\overline{B}} &= (\overline{AB})(\overline{A\overline{B}}) = (\overline{A}+\overline{B})(A+B) \\
 &= (\overline{A}+\overline{B})A + (\overline{A}+\overline{B})B = \overline{A}A + \overline{B}A + \overline{A}B + \overline{B}B \\
 &= \overline{A}\overline{B} + \overline{B}A = (A-B)(B-A).
 \end{aligned}$$

(4)因为 $A-B = \overline{A\overline{B}}$, $A = AB + A\overline{B}$, 而 AB 与 $A\overline{B}$ 互不相容, 所以

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

即 $P(A-B) = P(A) - P(AB).$

注 特别地, 当 $B \subset A$ 时, $A\overline{B} = \emptyset$, 则有 $P(A-B) = P(A) - P(B).$

2. 利用概率的加法公式计算随机事件的概率

例 1.2.5 某城市的电话号码由 7 位数组成(第一位数不为 0), 求下列事件的概率：

(1) A = “电话号码中不含 0 或 9”;

(2) B = “电话号码中含 0 但不含 9”.

(1)解法 1 利用古典概型公式计算.

基本事件总数 $n = 9 \times 10^6$, 事件 A 包含的基本事件数 $m = 9^7 + 8 \times 9^6 - 8^7$, 所以

$$P(A) = \frac{m}{n} = \frac{9^7 + 8 \times 9^6 - 8^7}{9 \times 10^6} = 0.77.$$

解法 2 利用加法公式计算.

设 H_1 = “电话号码中不含 0”, H_2 = “电话号码中不含 9”, 而

$$P(H_1) = \frac{9^7}{9 \times 10^6}, P(H_2) = \frac{8 \times 9^6}{9 \times 10^6}, P(H_1 H_2) = \frac{8^7}{9 \times 10^6},$$

由加法公式有

$$\begin{aligned} P_6(A) &= P_7(H_1 + H_2) = P(H_1) + P(H_2) - P(H_1 H_2) \\ &= \frac{9^7}{9 \times 10^6} + \frac{8 \times 9^6}{9 \times 10^6} - \frac{8^7}{9 \times 10^6} = 0.77. \end{aligned}$$

注 两种方法所得结果完全相同,但利用加法公式比直接计算显然更容易理解.

(2)解 基本事件总数 $n = 9 \times 10^6$, 事件 B 包含的基本事件数 $m = 8 \times 9^6 - 8^7$, 所以

$$P(B) = \frac{m}{n} = \frac{8 \times 9^6 - 8^7}{9 \times 10^6} = 0.24.$$

例 1.2.6 将编号为 1 2 3 的三本书任意地排列在书架上,求至少有一本书自左至右地排列顺序号与它的编号相同的概率.

解 设 $A =$ “至少有一本书自左至右地排列顺序号与它的编号相同”, $H_i =$ “第 i 本书的顺序号与它的编号相同”, $i = 1, 2, 3$.

显然 $A = H_1 + H_2 + H_3$, 其中 H_1, H_2, H_3 相容.

$$P(H_1) = P(H_2) = P(H_3) = \frac{2!}{3!} = \frac{1}{3};$$

$$P(H_1 H_2) = P(H_2 H_3) = P(H_1 H_3) = \frac{1}{3!} = \frac{1}{6};$$

$$P(H_1 H_2 H_3) = \frac{1}{3!} = \frac{1}{6}.$$

所以 $P(A) = P(H_1 + H_2 + H_3)$

$$\begin{aligned} &= P(H_1) + P(H_2) + P(H_3) - P(H_1 H_2) - P(H_2 H_3) - P(H_1 H_3) + \\ &P(H_1 H_2 H_3) \\ &= \frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \frac{1}{6} - \frac{1}{6} - \frac{1}{6} + \frac{1}{6} = \frac{2}{3}. \end{aligned}$$

例 1.2.7 袋中有 a 个白球和 b 个黑球,从中任取 n 个球($a + b \geq n$),求至少取出一个白球的概率.

解法 1 设 $A =$ “取出的 n 个球中至少有一个白球”, $H_i =$ “取出的 n 个球中恰好有 i 个白球”, $i = 1, 2, \dots, n$. 则 $H_1, H_2, \dots, H_n$ 互不相容,并且 $A = H_1 + H_2 + \dots + H_n$, 而 $P(H_i) = \frac{C_a^i C_b^{n-i}}{C_{a+b}^n}$, $i = 1, 2, \dots, n$. 则

$$\begin{aligned} P(A) &= P(H_1) + P(H_2) + \dots + P(H_n) = \frac{C_a^1 C_b^{n-1}}{C_{a+b}^n} + \frac{C_a^2 C_b^{n-2}}{C_{a+b}^n} + \dots + \frac{C_a^n C_b^0}{C_{a+b}^n} \\ &= \frac{1}{C_{a+b}^n} \sum_{i=1}^n C_a^i C_b^{n-i}. \end{aligned}$$

解法 2 设 $\bar{A} =$ “取出的 n 个球全是黑球”, 则 $P(\bar{A}) = \frac{C_b^n}{C_{a+b}^n}$, 所以

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{C_b^n}{C_{a+b}^n}.$$

注 由此题还可得到恒等式 $\sum_{i=0}^n C_a^i C_b^{n-i} = C_{a+b}^n$.

例 1.2.8 袋中装有 $N-1$ 只黑球及 1 只白球,每次从袋中随机地摸出一球,并换入一只黑球,这样继续下去,问第 k 次摸球时,摸到黑球的概率是多少?

解 设 A = “第 k 次摸球时摸到一只黑球”,则 $\bar{A}$ = “第 k 次摸球时摸到一只白球”.

因为袋中只有一只白球,而每次摸出一球后总是换入一只黑球,所以为了保证第 k 次摸到白球,前面的 $k-1$ 次一定不能摸到白球,从而 $\bar{A}$ 等价于事件“前 $k-1$ 次摸球每次都摸到黑球,第 k 次摸到白球”,因此 $P(\bar{A}) = \frac{(N-1)^{k-1} \times 1}{N^k} = \left(1 - \frac{1}{N}\right)^{k-1} \cdot \frac{1}{N}$, 所以

$$P(A) = 1 - P(\bar{A}) = 1 - \left(1 - \frac{1}{N}\right)^{k-1} \cdot \frac{1}{N}.$$

例 1.2.9 掷硬币 $2n$ 次,求出现正面的次数多于出现反面次数的概率.

解 设 H_1 = “出现正面次数多于反面次数”, H_2 = “出现反面次数多于正面次数”, H_3 = “出现正面次数等于反面次数”. 显然 $H_1 + H_2 + H_3 = \Omega$ 并且 H_1, H_2, H_3 互不相容,因此,有

$$P(H_1 + H_2 + H_3) = P(H_1) + P(H_2) + P(H_3) = P(\Omega) = 1.$$

由于正、反面地位是对称的(即正、反面是可以互换的),因此 $P(H_1) = P(H_2)$, 所以 $P(H_1)$

$$= \frac{1 - P(H_3)}{2}, \text{ 而 } P(H_3) = \frac{C_{2n}^n}{2^{2n}}, \text{ 故 } P(H_1) = \frac{1 - \frac{C_{2n}^n}{2^{2n}}}{2}.$$

注 在古典概型问题中利用事件的对称性,往往有较好的效果.

同步练习 1.2

1. 某人投篮两次,设事件 H_1 = “第一次投中”, H_2 = “第二次投中”,用 H_1 与 H_2 表示下列事件:

A = “两次都投中”, B = “两次都未投中”, C = “恰有一次投中”, D = “至少有一次投中”;并指出 A, B, C, D 哪些是互不相容事件,哪些是对立事件?

2. 化简下列各式:

(1) $\overline{AB} + \overline{AB} + \overline{AB} + AB$;

(2) $B + \overline{(\bar{A} + B)A}$;

(3) $ABC + \overline{ABC} + \overline{ABC} + \overline{ABC} + \overline{ABC} + \overline{ABC} + ABC$;

(4) $(A + B)(A + \bar{B})(\bar{A} + B)(\bar{A} + \bar{B})$.

3. 设事件 A, B 为两个任意事件,证明:

$$1 - P(\bar{A}) - P(\bar{B}) \leq P(AB) \leq P(A + B) \leq P(A) + P(B).$$

4. 假设接收一批产品时,检验其中的一半,若不合格品不超过 2%,则接收,否则拒收.现有 100 件产品,其中有 5 件不合格品,求该批产品被接收的概率.

5. 在 1 500 个产品中有 400 个次品,现从中任取 200 个,求下列事件的概率:

A = “恰有 90 个次品”, B = “至少有 2 个次品”.

6. 从 1~2 000 中任意取一整数,求取到的整数不能被 6 或 8 整除的概率.

7. 某班级有 50 人,问至少有两个人的生日在同一天的概率是多少?

8. 向 3 个相邻的军火库投一颗炸弹,炸中第一个军火库的概率为 0.025,炸中其余两个军火库的概率均为 0.1,只要炸中一个,另外两个也要发生爆炸,求军火库发生爆炸的概率.

9. 设 10 个球中有 2 个红球,现任取 n 个球,为了使这 n 个球中至少有一个红球的概率大于 0.5,问 n 至少应是多大?

同步练习 1.2 参考答案

$$1. A = H_1 H_2, B = \bar{H}_1 \bar{H}_2, C = H_1 \bar{H}_2 + \bar{H}_1 H_2, D = H_1 + H_2.$$

A, B, C 两两互不相容, B 和 D 对立.

$$2. (1)\Omega; (2)\Omega; (3)A+B+C; (4)\emptyset.$$

$$3. (1) \text{因为 } P(A+B) = P(A) + P(B) - P(AB), P(AB) \geq 0,$$

所以 $P(A+B) \leq P(A) + P(B)$;

$$(2) \text{因为 } AB \subset A+B, \text{所以 } P(AB) \leq P(A+B);$$

$$(3) 1 - P(\bar{A}) - P(\bar{B}) = P(A) + P(B) - 1 = P(AB) + P(A+B) - 1 \\ = P(AB) - (1 - P(A+B)) \leq P(AB).$$

$$4. 0.1810.$$

$$5. P(A) = \frac{C_{400}^{90} \cdot C_{1100}^{110}}{C_{1500}^{200}}, P(B) = 1 - \frac{C_{400}^1 \cdot C_{1100}^{199}}{C_{1500}^{200}} - \frac{C_{1100}^{200}}{C_{1500}^{200}}.$$

$$6. \frac{3}{4}.$$

$$7. 1 - \frac{C_{365}^{50} \cdot 50!}{365^{50}}. 8. 0.225. 9. n \geq 3.$$

1.3 条件概率与事件的独立性

1.3.1 知识要点

1. 条件概率

如果 A 及 B 是条件组 S 下的两个随机事件,且 $P(A) \neq 0$,则称在事件 A 发生前提下事件 B 发生的概率为条件概率,记为 $P(B|A)$.

$$P(B|A) = \frac{P(AB)}{P(A)} \quad (P(A) \neq 0).$$

条件概率是概率论中的一个重要概念,条件概率也是概率,它具有概率的一切性质.

2. 乘法公式

①对于任意两个事件 A 与 B , $P(A) > 0$, $P(B) > 0$,有

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

②对于任意 n 个事件 $A_1, A_2, \dots, A_n$ 有

$$P(A_1 A_2 \dots A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 A_2) \cdot \dots \cdot P(A_n | A_1 A_2 \dots A_{n-1}).$$

3. 独立性

①如果事件 A 及 B 满足 $P(AB) = P(A)P(B)$ 则称 A 与 B 为相互独立的随机事件.

注 1 两事件相互独立的含义是它们中的一个已发生, 不影响另一个发生的概率.

注 2 如果事件 A 与 B 独立, 且 $P(A) > 0$ 则 $P(B|A) = P(B)$, 反之亦然.

②如果事件 A 与 B , A 与 $\bar{B}$, $\bar{A}$ 与 B , $\bar{A}$ 与 $\bar{B}$ 中有一对相互独立, 则另外三对也相互独立.

③如果事件 A, B, C 满足 $P(AB) = P(A)P(B)$, $P(BC) = P(B)P(C)$, $P(CA) = P(C)P(A)$, $P(ABC) = P(A)P(B)P(C)$ 则称 A, B, C 是相互独立的事件.

④如果事件组 $A_1, A_2, \dots, A_n$ 对于任意的正整数 $k (2 \leq k \leq n)$ 和任意的 $1 \leq i_1 < i_2 < \dots < i_k \leq n$ 均满足等式 $P(A_{i_1} A_{i_2} \dots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$ 则称事件 $A_1, A_2, \dots, A_n$ 是相互独立的.

注 ①如果事件 $A_1, A_2, \dots, A_n (n \geq 2)$ 相互独立, 则其中任意 $k (2 \leq k \leq n)$ 个事件也是相互独立的.

②如果事件 $A_1, A_2, \dots, A_n (n \geq 2)$ 相互独立, 则将其中任意多个事件换成它们的对立事件, 所得的 n 个事件仍相互独立.

1.3.2 典型题型解析

1. 计算随机事件的条件概率

例 1.3.1 验收 100 只电器元件品, 随意抽验 5 件, 如果全部为合格品则接收, 否则拒收. 假如 100 只电器元件中恰有 5 件不合格品, 求这批产品经检验被拒收的概率.

解 设 A = “检验后产品被拒收”, H_i = 抽验的第 i 件是合格品 ($i = 1, 2, \dots, 5$). 则 $\bar{A}$ = “抽验的 5 件产品都是合格品”, 显然 $\bar{A} = H_1 H_2 H_3 H_4 H_5$, $P(\bar{A}) = P(H_1 H_2 H_3 H_4 H_5) = P(H_1)P(H_2|H_1)P(H_3|H_1 H_2)P(H_4|H_1 H_2 H_3)P(H_5|H_1 H_2 H_3 H_4) = \frac{95}{100} \cdot \frac{94}{99} \cdot \frac{93}{98} \cdot \frac{92}{97} \cdot \frac{91}{96} = 0.77$, 所以 $P(A) = 1 - P(\bar{A}) = 0.23$.

例 1.3.2 某地区气象资料表明, 邻近的甲、乙两城市中全年雨天的比例: 甲市为 12%, 乙市为 10%, 甲、乙两市至少有一市为雨天的比例为 16%. 求:

- (1) 甲、乙两市同为雨天的概率;
- (2) 甲市雨天的条件下, 乙市也为雨天的概率;
- (3) 甲市无雨的条件下, 乙市也无雨的概率.

解 设 A = “甲市为雨天”, B = “乙市为雨天”. 根据已知条件有 $P(A) = 0.12$, $P(B) = 0.1$, $P(A + B) = 0.16$ 则

$$(1) P(AB) = P(A) + P(B) - P(A + B) = 0.12 + 0.1 - 0.16 = 0.06;$$

$$(2) P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.06}{0.12} = 0.5;$$

$$(3) P(\bar{B}|\bar{A}) = \frac{P(\bar{A}\bar{B})}{P(\bar{A})} = \frac{P(\overline{A+B})}{P(\bar{A})} = \frac{1 - P(A+B)}{1 - P(A)} = \frac{1 - 0.16}{1 - 0.12} = 0.95.$$

例 1.3.3 设 10 件产品中有 4 件不合格品, 从中任取两件, 已知在所取的两件产品中至

少有一件是不合格品,求另一件也是不合格品的概率.

解 设 A = “两件产品中至少有一件是不合格品”,

B = “两件产品都是不合格品”,

C = “两件产品中一件是合格品,一件是不合格品”

$A = B + C$ 并且 $BC = \emptyset$, 所以 $P(A) = P(B + C) = P(B) + P(C) = \frac{C_4^2}{C_{10}^2} + \frac{C_4^1 \cdot C_6^1}{C_{10}^2} = \frac{2}{3}$, 由

于 $B \subset A$, 所以 $AB = B$ 则 $P(AB) = P(B) = \frac{C_4^2}{C_{10}^2} = \frac{2}{15}$, 所以根据条件概率公式所求概率为

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{2}{15}}{\frac{2}{3}} = \frac{1}{5}.$$

例 1.3.4 在空战中,甲机先向乙机开火,击落乙机的概率为 0.2,若乙机未被击中,进行还击,击落甲机的概率是 0.3,若甲机未被击落,再次进攻乙机,击落乙机的概率是 0.4,求在这 3 次攻击中,甲、乙机各被击中的概率.

解 A_i = “甲机第 i 次射击成功” ($i = 1, 2$), B_1 = “乙机射击成功”, A = “甲机被击落”, B = “乙机被击落”.

$$P(A) = P(\bar{A}_1 B_1) = P(\bar{A}_1)P(B_1|\bar{A}_1) = 0.8 \times 0.3 = 0.24;$$

$$P(B) = P(A_1 + \bar{A}_1 \bar{B}_1 A_2) = P(A_1) + P(\bar{A}_1 \bar{B}_1 A_2) = P(A_1) +$$

$$P(\bar{A}_1)P(\bar{B}_1|\bar{A}_1)P(A_2|\bar{A}_1 \bar{B}_1) = 0.2 + 0.8 \times 0.7 \times 0.4 = 0.424.$$

例 1.3.5 在 1 2 3 4 5 这五个数码中,每次取一个数码,取后不放回,连续取两次,求在第一次取到偶数的条件下,第二次取到奇数的概率.

解 设 A = “第一次取出偶数”, B = “第二次取出奇数”.

解法 1 根据已知条件可求得基本事件总数 $n = A_5^2$, 基本事件为 (1, 2), (1, 3), (1, 4), (1, 5), (2, 1), (2, 3), (2, 4), (2, 5), (3, 1), (3, 2), (3, 4), (3, 5), (4, 1), (4, 2), (4, 3), (4, 5), (5, 1), (5, 2), (5, 3), (5, 4). 而 $P(A) = \frac{C_2^1 \cdot C_4^1}{A_5^2} = \frac{2}{5}$, $P(AB) = \frac{C_2^1 \cdot C_3^1}{A_5^2} = \frac{3}{10}$, 所以由条件概率公式可得

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{3}{10}}{\frac{2}{5}} = \frac{3}{4}.$$

解法 2 把第一次取到偶数,即事件 A 的发生条件作为随机试验的先决条件,这时基本事件为 (2, 1), (2, 3), (2, 4), (2, 5), (4, 1), (4, 2), (4, 3), (4, 5), 由这些基本事件组成的基本事件组称之为事件 B 的“缩减基本事件组”. 基本事件总数 $n = C_2^1 \cdot C_4^1 = 8$, 其中第二个数码是奇数的基本事件数 $m = C_2^1 \cdot C_3^1$, 所以由古典概型公式可得 $P(B|A) = \frac{m}{n} = \frac{C_2^1 \cdot C_3^1}{C_2^1 \cdot C_4^1} = \frac{6}{8} = \frac{3}{4}$.

注 解法 2 中,将事件 A 的发生作为先决条件,确定事件 B 的缩减基本事件组,然后利用古典概型公式直接计算条件概率 $P(B|A)$, 我们不妨称之为“缩减基本事件法”.

例 1.3.6 在一个盒子中混有新、旧两种乒乓球. 新乒乓球中有白色球 40 个, 红色球 30

个,旧乒乓球中有白色球 20 个,红色球 10 个,在这个盒子中任取一球,发现是新的,求这个乒乓球是白色的概率.

解 设 A = “取到新球”, B = “取到白色球”.

$$\text{解法 1 } P(A) = \frac{7}{10}, P(AB) = \frac{2}{5}, \text{ 所以 } P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{2}{5}}{\frac{7}{10}} = \frac{4}{7}.$$

解法 2 将事件 A 即取到新球作为先决条件,利用“缩减基本事件法”考虑,由于在 70 个新球中共有 40 个白色球,所以 $P(B|A) = \frac{4}{7}$.

2. 利用事件的独立性计算概率

例 1.3.7 5 个乒乓球,其中 3 个旧球,2 个新球,每次取 1 个,连续取 2 次,求至少有一次取到新球的概率.

解 设 A_i = “第 i 次取到新球” ($i = 1, 2$); B = “至少有 1 次取到新球”, $B = A_1 + A_2$.

(1) 有放回地取球.

解法 1 由于两次取球相互独立,所以事件 A_1 与 A_2 相互独立,即

$$P(B) = P(A_1 + A_2) = 1 - P(\bar{A}_1 \bar{A}_2) = 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) = 1 - \frac{3}{5} \cdot \frac{3}{5} = \frac{16}{25}.$$

解法 2 $P(B) = P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$

$$= P(A_1) + P(A_2) - P(A_1)P(A_2) = \frac{2}{5} + \frac{2}{5} - \frac{2}{5} \cdot \frac{2}{5} = \frac{16}{25}.$$

(2) 无放回地取球.

解法 1 由于是无放回取球,所以事件 A_1 与 A_2 不独立,即

$$\begin{aligned} P(B) &= P(A_1 + A_2) = 1 - P(\bar{A}_1 \bar{A}_2) = 1 - P(\bar{A}_1)P(\bar{A}_2 | \bar{A}_1) \\ &= 1 - \frac{3}{5} \cdot \frac{2}{4} = \frac{7}{10}. \end{aligned}$$

解法 2 $P(B) = P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2)$

$$= P(A_1) + P(A_2) - P(A_1)P(A_2 | A_1) = \frac{2}{5} + \frac{C_2^1 \cdot C_4^1}{C_5^1 \cdot C_4^1} - \frac{2}{5} \cdot \frac{1}{4} = \frac{7}{10}.$$

注 求事件 $A_1, A_2, \dots, A_n$ 的和的概率虽然有加法公式,但当事件数目较大或事件相容时,直接利用此公式计算一般较困难,为此,常用对立事件的概率公式求之,因为 n 个事件的和的对立事件等于这 n 个事件的对立事件的积,从而将和的概率转化为积的概率,即

$$P(A_1 + A_2 + \dots + A_n) = 1 - P(\bar{A}_1 + \bar{A}_2 + \dots + \bar{A}_n) = 1 - P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_n).$$

特别当 $A_1, A_2, \dots, A_n$ 相互独立时,有 $P(A_1 + A_2 + \dots + A_n) = 1 - P(\bar{A}_1)P(\bar{A}_2) \dots P(\bar{A}_n)$.

例 1.3.8 设 A, B 为两个事件,且 $P(A|B) = P(A|\bar{B})$, $P(A) > 0$, $P(B) > 0$,证明: A 与 B 独立.

证明 因为 $P(A|B) = \frac{P(AB)}{P(B)}$, $P(A|\bar{B}) = \frac{P(A\bar{B})}{P(\bar{B})}$, 又已知 $P(A|B) = P(A|\bar{B})$, 所以

$\frac{P(AB)}{P(B)} = \frac{P(A\bar{B})}{P(\bar{B})}$, 又因为 $A = A(B + \bar{B}) = AB + A\bar{B}$, 并且 AB 与 $A\bar{B}$ 互不相容, 所以 $P(A) = P(AB) + P(A\bar{B})$, 即 $P(AB) = P(A) - P(A\bar{B})$, 因此,

$P(AB)(1-P(B))=(P(A)-P(AB))P(B)$, 即 $P(AB)=P(A) \cdot P(B)$,
所以事件 A 与 B 互相独立.

同步练习 1.3

1. 10 个零件中有 3 个次品, 每次从其中任取一个零件, 取后不放回, 求第 3 次才取得合格品的概率.

2. 某地区位于河流甲与河流乙的汇合点, 当任一河流泛滥时, 该地区即被淹没. 设在某时期内河流甲泛滥的概率是 0.1, 河流乙泛滥的概率是 0.2, 当河流甲泛滥时引起河流乙泛滥的概率为 0.3, 求在该时期内这个地区被淹没的概率及当河流乙泛滥时引起河流甲泛滥的概率.

3. 甲乙两人独立地对同一目标射击一次, 其命中率分别为 0.6 和 0.5, 现已知目标被击中, 则它是甲射中的概率是多大?

4. 一盒子中有 4 只次品晶体管, 6 只正品晶体管, 逐个抽取进行测试, 测试后不放回, 直到 4 只次品晶体管都被找到为止. 求第 4 只次品晶体管在 (1) 第 5 次测试时发现; (2) 第 10 次测试时被发现的概率.

5. 轰炸机要完成它的使命, 必须是驾驶员找到了目标, 同时投弹员投中了目标. 设驾驶员甲和乙找到目标的概率分别为 0.9 和 0.8, 投弹员丙和丁在驾驶员找到目标的前提下投中目标的概率分别为 0.7 和 0.6. 现在要配备两架轰炸机的人员, 问甲、乙、丙、丁应怎样两两组合才使完成任务的概率较大 (注意, 只要有一架飞机投中目标即完成任务)?

6. 据资料表明, 某三口之家患某种传染病的概率有以下规律: 孩子得病的概率为 0.6, 在孩子得病的条件下母亲也得病的概率为 0.5, 在孩子与母亲都得病的前提下, 父亲也得病的概率为 0.4, 求母亲与孩子得病而父亲未得病的概率.

7. 某人忘记了电话号码的最后一个数字, 因而他随意地拨号, 求他拨号不超过 3 次而接通所需电话的概率.

8. 3 人独立地去破译一份密码, 已知各人能译出的概率分别为 $\frac{1}{5}, \frac{1}{3}, \frac{1}{4}$. 求 3 人中至少有 1 人能译出密码的概率.

9. 设第 1 只盒子中装有 3 只蓝球, 2 只绿球, 2 只白球, 第 2 只盒子中装有 2 只蓝球, 3 只绿球, 4 只白球, 独立地分别从两只盒子中各取 1 只球, 求:

(1) 至少有 1 只蓝球的概率;

(2) 有 1 只蓝球 1 只白球的概率;

(3) 已知至少有 1 只蓝球的前提下, 有 1 只蓝球和 1 只白球的概率.

10. 假设参加比赛的 15 名选手中 5 名是种子选手, 现将 15 人随意分成 5 组, 求每组各有 1 名种子选手的概率.

11. 掷三颗骰子, 已知得到的三个点数不同, 求其中已含有 1 点的概率.

12. 袋中装有 $2n-1$ 个白球, $2n$ 个黑球, 一次取出 n 个球, 发现都是同一种颜色的, 求这种颜色是黑色的概率.

13. 设事件 A 与 B 独立, 并且 $P(A\bar{B})=P(\bar{A}B)=\frac{1}{4}$, 求 $P(A)$ 与 $P(B)$.

14. 已知 $P(\bar{A})=0.3, P(B)=0.4, P(A\bar{B})=0.5$, 求 $P(B|A \cup \bar{B})$.

15. 设 $P(A)>0, P(B)>0, P(A|B)+P(\bar{A}|B)=1$, 证明: 事件 A 与 B 互相独立.

16. 设事件 A 与事件 B_1 和 B_2 互相独立, 并且 B_1 和 B_2 互不相容, 证明 A 与 B_1+B_2 互相独立.

同步练习 1.3 参考答案

$$1. 0.058, 2. 0.27, 3. 0.15, 4. \frac{2}{105}, \frac{2}{5}.$$

5. 甲丙、乙丁组合完成任务的概率为 0.807 6, 甲丁、乙丙组合完成的概率为 0.797 6.

$$6. 0.18, 7. \frac{3}{10}, 8. 0.6, 9. \frac{5}{9}, \frac{16}{63}, \frac{16}{35}, 10. 0.080, 11. \frac{1}{2}, 12. \frac{2}{3}.$$

$$13. P(A) = P(B) = \frac{1}{2}, 14. 0.25.$$

1.4 全概率公式

1.4.1 知识要点

1. 完备事件组

如果事件组 $A_1, A_2, \dots, A_n$ 满足

(1) $A_1, A_2, \dots, A_n$ 互不相容, 且 $P(A_i) > 0 (i = 1, 2, \dots, n)$,

(2) $A_1 + A_2 + \dots + A_n = \Omega$,

则称事件组 $A_1, A_2, \dots, A_n$ 为完备事件组.

2. 全概率公式

如果事件组 $A_1, A_2, \dots, A_n$ 是一完备事件组, 则对任一事件 B 均有

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

全概率公式是应用非常广泛的一个公式, 它把事件 B 的概率(不易计算)分解成 n 个比较容易计算的概率之和, 形似繁难, 实则简单.

用全概率公式求事件 B 的概率 $P(B)$ 的关键是找一个完备事件组 $A_1, A_2, \dots, A_n$. 事件 B 必伴随着 $A_1, A_2, \dots, A_n$ 之一发生.

3. Bayes 公式

如果 $A_1, A_2, \dots, A_n$ 是一完备事件组, 则对任一事件 $B (P(B) \neq 0)$ 均有

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)} \quad (j = 1, 2, \dots, n).$$

Bayes 公式也称为逆概公式, Bayes 公式实际上是计算条件概率, 它是在已知结果发生的情况下, 求导致结果的某种原因的概率. Bayes 公式是条件概率与全概率公式的综合运用.

4. Bernoulli 试验

设在单次试验中, 只有两个可能的结果 A 及 $\bar{A}$, 且 $P(A) = p (0 < p < 1)$. 如果将此试验完全独立地重复 n 次, 则称这一串试验为 Bernoulli 试验或 n 重独立试验.

5. Bernoulli 概型公式

设在单次试验中,事件 A 发生的概率为 $p(0 < p < 1)$,则在 n 次独立试验中事件 A 恰好发生 k 次的概率为 $C_n^k p^k (1-p)^{n-k}$,即如果令 $A_k = \text{"}A \text{ 恰好发生 } k \text{ 次"}$, $q = 1 - p$,则有

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

上述公式称为 Bernoulli 概型公式, Bernoulli 概型与古典概型不同,虽然基本事件的总数是有限的(共 2^n 个),但这些基本事件不一定是等概的.

1.4.2 典型题型解析

1. 利用全概率公式和 Bayes 公式计算概率

例 1.4.1 有朋友自远方来访,他乘火车、轮船、汽车、飞机来的概率分别是 0.3, 0.2, 0.1, 0.4. 如果他乘火车、轮船、汽车来的话,迟到的概率分别是 $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{12}$, 而乘飞机则不会迟到,结果朋友迟到了,求他乘火车来的概率.

解 设 $A_1 = \text{"朋友乘火车"}$, $A_2 = \text{"朋友乘轮船"}$, $A_3 = \text{"朋友乘汽车"}$, $A_4 = \text{"朋友乘飞机"}$, $B = \text{"朋友迟到"}$. 则根据已知条件可知 $P(A_1) = 0.3$, $P(A_2) = 0.2$, $P(A_3) = 0.1$, $P(A_4) = 0.4$, $P(B|A_1) = \frac{1}{4}$, $P(B|A_2) = \frac{1}{3}$, $P(B|A_3) = \frac{1}{12}$, $P(B|A_4) = 0$, 由全概公式可得

$$P(B) = \sum_{i=1}^4 P(A_i)P(B|A_i) = 0.3 \times \frac{1}{4} + 0.2 \times \frac{1}{3} + 0.1 \times \frac{1}{12} + 0.4 \times 0 = \frac{3}{20}.$$

由 Bayes 公式可得

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{0.3 \times 0.25}{\frac{3}{20}} = \frac{1}{2}.$$

例 1.4.2 设在同一年级有两个班,一班 50 名学生,其中 10 名女生;二班 30 名学生,其中 18 名女生.在两班中任选一班,然后从中先后挑选两名学生,求:

(1)先选出的是女生的概率;

(2)在已知选出的是女生的条件下,后选出的学生也是女生的概率.

解 设 $B_i = \text{"被选到的是第 } i \text{ 班"}$ ($i=1, 2$),

$A_j = \text{"第 } j \text{ 次选出的是女生"}$ ($j=1, 2$).

由题意可得 $P(B_1) = \frac{1}{2}$, $P(B_2) = \frac{1}{2}$, $P(A_1|B_1) = \frac{1}{5}$, $P(A_1|B_2) = \frac{3}{5}$.

$$(1) P(A_1) = \sum_{i=1}^2 P(B_i)P(A_1|B_i) = \frac{1}{2} \cdot \frac{1}{5} + \frac{1}{2} \cdot \frac{3}{5} = \frac{2}{5}.$$

$$(2) P(A_1 A_2) = \sum_{i=1}^2 P(B_i)P(A_1 A_2|B_i) = \frac{1}{2} \cdot \frac{A_{10}^2}{A_{50}^2} + \frac{1}{2} \cdot \frac{A_{18}^2}{A_{30}^2} \\ = \frac{1}{2} \cdot \frac{1}{5} \cdot \frac{9}{49} + \frac{1}{2} \cdot \frac{3}{5} \cdot \frac{17}{29} = 0.194.$$

其中 $P(A_1 A_2|B_1)$ 表示在选中一班的情况下两次都挑出女生的概率, $P(A_1 A_2|B_2)$ 表示在选中二班的情况下两次都挑出女生的概率. 所以 $P(A_2|A_1) = \frac{P(A_1 A_2)}{P(A_1)} = \frac{0.194}{0.4} = 0.485$.

例 1.4.3 玻璃杯成箱出售,每箱 20 只,假设各箱含 0, 1, 2 只残次品的概率分别为 0.8, 0.1, 0.1. 一顾客欲购一箱玻璃杯,售货员随机地取一箱,顾客开箱随机地察看 4 只,若无残次

品,则买下该箱玻璃杯,否则退回.求:

(1)顾客买下该箱的概率;

(2)在顾客买下的一箱中,确实没有残次品的概率.

解 设 B = “顾客买下该箱”, A_i = “该箱有 i 件残次品” ($i = 0, 1, 2$), 由题意知 $P(A_0) = 0.8$, $P(A_1) = 0.1$, $P(A_2) = 0.1$.

$$\begin{aligned}(1) P(B) &= \sum_{i=0}^2 P(A_i)P(B|A_i) = 0.8 \times \frac{C_{20}^4}{C_{20}^4} + 0.1 \times \frac{C_{19}^4}{C_{20}^4} + 0.1 \times \frac{C_{18}^4}{C_{20}^4} \\ &= 0.8 \times 1 + 0.1 \times \frac{4}{5} + 0.1 \times \frac{12}{19} = 0.94.\end{aligned}$$

$$(2) P(A_0|B) = \frac{P(A_0)P(B|A_0)}{P(B)} = \frac{0.8 \times 1}{0.94} = 0.85.$$

例 1.4.4 已知某城市下雨的时间占一半,天气预报的准确性为 0.9,某人每天上班预报下雨他就带伞,即使预报无雨,他也有一半时间带伞.求:

(1)已知他没带伞,却遇到下雨的概率;

(2)已知他带伞,但天不下雨的概率.

解 设 A = “天下雨”, B = “预报下雨”, C = “某人带伞”. 由题意知

$$P(A) = P(\bar{A}) = \frac{1}{2}, P(C|AB) = P(C|\bar{A}B) = 1, P(C|A\bar{B}) = P(C|\bar{A}\bar{B}) = \frac{1}{2},$$

$$P(B|A) = P(\bar{B}|\bar{A}) = 0.9, P(\bar{B}|A) = P(B|\bar{A}) = 0.1.$$

而 $P(C|A) = P(C|BA)P(B|A) + P(C|\bar{B}A)P(\bar{B}|A)$

$$= 1 \times 0.9 + \frac{1}{2} \times 0.1 = \frac{19}{20},$$

$$P(C|\bar{A}) = P(C|\bar{A}B)P(B|\bar{A}) + P(C|\bar{A}\bar{B})P(\bar{B}|\bar{A})$$

$$= 1 \times 0.1 + \frac{1}{2} \times 0.9 = \frac{11}{20},$$

$$P(\bar{C}|A) = 1 - P(C|A) = \frac{1}{20},$$

$$P(\bar{C}|\bar{A}) = 1 - P(C|\bar{A}) = \frac{9}{20}.$$

$$(1) P(A|\bar{C}) = \frac{P(A) \cdot P(\bar{C}|A)}{P(A)P(\bar{C}|A) + P(\bar{A})P(\bar{C}|\bar{A})} = \frac{\frac{1}{2} \times \frac{1}{20}}{\frac{1}{2} \times \frac{1}{20} + \frac{1}{2} \times \frac{9}{20}} = \frac{1}{10}.$$

$$(2) P(\bar{A}|C) = \frac{P(\bar{A})P(C|\bar{A})}{P(A)P(C|A) + P(\bar{A})P(C|\bar{A})} = \frac{\frac{1}{2} \times \frac{11}{20}}{\frac{1}{2} \times \frac{19}{20} + \frac{1}{2} \times \frac{11}{20}} = \frac{11}{30}.$$

例 1.4.5 设有来自 3 个地区的各 10 名, 15 名和 25 名考生的报名表, 其中女生的报名表分别为 3 份, 7 份和 5 份, 随机地取 1 个地区的报名表, 从中先后抽出两份.

(1)求先抽到的一份表是女生表的概率.

(2)已知后抽到的一份是男生表, 求先抽到的一份是女生表的概率.

解 设 H_i = “报名表是取自第 i 个地区” ($i = 1, 2, 3$);

A_j = “第 j 次取出的报名表是女生表” ($j = 1, 2$).

由题意可知 $P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$, $P(A_1 | H_1) = \frac{3}{10}$, $P(A_1 | H_2) = \frac{7}{15}$, $P(A_1 | H_3) = \frac{1}{5}$.

$$(1) P(A_1) = \sum_{i=1}^3 P(H_i)P(A_1 | H_i) = \frac{1}{3} \times \frac{3}{10} + \frac{1}{3} \times \frac{7}{15} + \frac{1}{3} \times \frac{1}{5} = \frac{29}{90}.$$

$$(2) P(A_1 | \bar{A}_2) = \frac{P(A_1 \bar{A}_2)}{P(\bar{A}_2)},$$

而 $P(\bar{A}_2 | H_1) = \frac{7}{10}$, $P(\bar{A}_2 | H_2) = \frac{8}{15}$, $P(\bar{A}_2 | H_3) = \frac{4}{5}$, $P(A_1 \bar{A}_2 | H_1) = \frac{3}{10} \times \frac{7}{9} = \frac{7}{30}$,

$$P(A_1 \bar{A}_2 | H_2) = \frac{7}{15} \times \frac{8}{14} = \frac{8}{30}, P(A_1 \bar{A}_2 | H_3) = \frac{1}{5} \times \frac{20}{24} = \frac{5}{30},$$

因此 $P(\bar{A}_2) = \sum_{i=1}^3 P(H_i)P(\bar{A}_2 | H_i) = \frac{1}{3} \times \frac{7}{10} + \frac{1}{3} \times \frac{8}{15} + \frac{1}{3} \times \frac{4}{5} = \frac{61}{90}$,

$$P(A_1 \bar{A}_2) = \sum_{i=1}^3 P(H_i)P(A_1 \bar{A}_2 | H_i) = \frac{1}{3} \times \frac{7}{30} + \frac{1}{3} \times \frac{8}{30} + \frac{1}{3} \times \frac{5}{30} = \frac{2}{90},$$

所以 $P(A_1 | \bar{A}_2) = \frac{P(A_1 \bar{A}_2)}{P(\bar{A}_2)} = \frac{\frac{2}{90}}{\frac{61}{90}} = \frac{2}{61}$.

2. 利用 Bernoulli 概型公式计算概率

例 1.4.6 甲、乙两名运动员进行乒乓球单打比赛, 已知每一局甲胜的概率为 0.6, 乙胜的概率为 0.4, 比赛时可以采用三局二胜制或五局三胜制, 问在哪一种赛制下, 甲获胜的可能性较大?

解 (1) 采用三局两胜制.

设 A_1 = “甲净胜两局”, A_2 = “前两局甲、乙各胜一局, 第三局甲胜”, A = “甲获胜”, 则 $A = A_1 + A_2$. 而

$$P(A_1) = C_2^2 (0.6)^2 (0.4)^0 = 0.36, P(A_2) = [C_2^1 (0.6)^1 (0.4)^1] \cdot 0.6 = 0.288.$$

所以 $P(A) = P(A_1 + A_2) = P(A_1) + P(A_2) = 0.36 + 0.288 = 0.648$.

(2) 采用五局三胜制.

设 B_1 = “甲净胜三局”, B_2 = “前三局中甲胜两局, 负一局, 第四局甲胜”, B_3 = “前四局甲、乙各胜两局, 第五局甲胜”, B_1, B_2, B_3 互不相容, 则 $A = B_1 + B_2 + B_3$, $P(A) = P(B_1 + B_2 + B_3) = P(B_1) + P(B_2) + P(B_3)$, 而 $P(B_1) = C_3^3 \cdot (0.6)^3 \cdot (0.4)^0 = 0.216$,

$$P(B_2) = [C_3^2 \cdot (0.6)^2 \cdot (0.4)^1] \times 0.6 = 0.259,$$

$$P(B_3) = [C_4^2 \cdot (0.6)^2 \cdot (0.4)^2] \times 0.6 = 0.207.$$

所以 $P(A) = 0.216 + 0.259 + 0.207 = 0.682$.

比较(1)、(2)的结果可知, 采用五局三胜的赛制, 甲获胜的可能性较大.

例 1.4.7 一射手对同一目标独立地进行 4 次射击, 若至少命中一次的概率为 $\frac{5}{6}$, 求该射手的命中率(即射击技术水平).

解 因为 n 重 Bernoulli 试验中事件 A 发生 0 次的概率为 $P(A_0) = C_n^0 \cdot p^0 \cdot (1-p)^n = (1-p)^n$, 所以事件 A 至少发生一次的概率为 $P(\text{“事件 } A \text{ 至少发生一次”}) = 1 - P(A_0) = 1 -$

$(1-p)^n$, 从而得到在一次试验中事件 A 发生的概率 p 的计算公式: $p = 1 - \sqrt[n]{1 - P(\text{“事件 } A \text{ 至少发生一次”})}$. 所以该射手的命中率为 $p = 1 - \sqrt[4]{1 - \frac{5}{6}} = 1 - \frac{1}{\sqrt[4]{6}} = 0.36$.

例 1.4.8 每次从 0~9 这 10 个数字中随机取一个, 取后放回, 得到的数列称为随机数字数列, 问随机数字数列要有多长才能使数字 0 至少出现一次的概率不小于 0.9?

解 在 Bernoulli 试验中, 设 $P(A) = p$, 为使事件 A 至少发生一次的概率不小于给定的 α , 又由例 7 可知事件 A 至少发生一次的概率为 $1 - (1-p)^n$, 因此, $1 - (1-p)^n \geq \alpha$ 即 $(1-p)^n \leq 1 - \alpha$, 解得 $n \geq \frac{\lg(1-\alpha)}{\lg(1-p)}$, 以 $[x]$ 表示不大于 x 的最大整数, 则 $n > \left\lceil \frac{\lg(1-\alpha)}{\lg(1-p)} \right\rceil$, 即在 Bernoulli 试验中, 为使事件 A 至少发生一次的概率不小于给定的 α , 至少需要做 $\left\lceil \frac{\lg(1-\alpha)}{\lg(1-p)} \right\rceil + 1$ 次试验.

在本题中设随机数字数列有 n 个数字, 把每取一个数字看作一次试验, 每次试验只有两种可能结果: $A = \text{“取到数 0”}$ 和 $\bar{A} = \text{“没有取到数 0”}$, 另外由于每次取数是有放回的, 所以, 该试验为 Bernoulli 试验. 由题意知 $P(A) = 0.1$, $\alpha = 0.9$, 为使数字 0 至少出现一次的概率不小于 0.9, 需取数的次数为 $n \geq \left\lceil \frac{\lg(1-0.9)}{\lg(1-0.1)} \right\rceil + 1 = 22$, 即随机数字数列最少要有 22 个数字.

例 1.4.9 掷一枚硬币, 一直到它出现正面为止, 求需要掷 r 次的概率.

解 进行一系列独立重复试验, 每次试验只有“成功”和“失败”两种结果, 设每次试验成功的概率为 p , 则直到第 r 次才成功意味着前 $r-1$ 次试验事件 A 都不发生, 第 r 次事件 A 发生, 所以, 由乘法公式可得第 r 次才成功的概率(记为 P_r)为

$$P_r = P(\underbrace{\bar{A}\bar{A}\dots\bar{A}}_{r-1\text{次}} A) = \underbrace{P(\bar{A})P(\bar{A})\dots P(\bar{A})}_{r-1\text{次}} P(A) = p(1-p)^{r-1}.$$

在本题中由于每次投掷硬币出现正面的概率 $p = \frac{1}{2}$, 所求概率为

$$P_r = \frac{1}{2} \left(1 - \frac{1}{2}\right)^{r-1} = \frac{1}{2^r}.$$

例 1.4.10 进行一系列独立重复试验, 每次试验只有“成功”和“失败”两种结果, 每次试验成功的概率为 p ($0 < p < 1$), 求证:

(1) 直到第 n 次试验才取得 r ($1 \leq r \leq n$) 次成功的概率为 $C_{n-1}^{r-1} p^r (1-p)^{n-r}$;

(2) 第 r 次成功之前恰失败 k 次的概率为 $C_{r+k-1}^{r-1} p^r (1-p)^k$.

解 (1) 设 $A = \text{“直到第 } n \text{ 次试验才取得 } r \text{ 次成功”}$, $B = \text{“前 } n-1 \text{ 次试验中恰有 } r-1 \text{ 次成功”}$, $C = \text{“第 } n \text{ 次试验成功”}$. 事件 A 发生意味着进行了 n 次独立试验, 前 $n-1$ 次中有 $r-1$ 次成功, 第 n 次成功, 即 $A = BC$, $P(A) = P(BC) = P(B) \cdot P(C)$. 由 Bernoulli 概型公式知 $P(B) = C_{n-1}^{r-1} \cdot p^{r-1} (1-p)^{n-r}$, 另外又知 $P(C) = p$, 所以

$$P(A) = P(B)P(C) = C_{n-1}^{r-1} \cdot p^{r-1} \cdot (1-p)^{n-r} \cdot p = C_{n-1}^{r-1} p^r (1-p)^{n-r}.$$

(2) 因为试验总次数为成功次数与失败次数之和, 所以, 试验总次数为 $r+k$, 从而本题是求重复进行试验直到第 $r+k$ 次才取得 r 次成功的概率, 所以利用(1)题的结果可得所求概率为 $C_{r+k-1}^{r-1} p^r (1-p)^k$.

例 1.4.11 (巴拿赫火柴盒问题) 某数学家随身带着两盒火柴, 每盒有 N 根火柴, 每次用时, 随机地任取一盒, 然后从中抽取一根. 求:

(1)发现一盒空而另一盒恰剩 r 根的概率是多少?

(2)在第一次用完一盒火柴(不是发现空)时另一盒恰剩 r 根的概率是多少?

解 (1) A = “发现一盒空而另一盒恰剩 r 根”, B = “发现甲盒空而乙盒恰剩 r 根”, C = “发现乙盒空而甲盒恰剩 r 根”, 则 $A = B + C$, 且 B 与 C 互不相容, 所以 $P(A) = P(B) + P(C)$, 下面求 $P(B)$ 与 $P(C)$. 如果该数学家首次发现甲盒是空的, 此时乙盒恰有 r 根, 那么此人取甲盒和乙盒的次数共有 $2N - r + 1$ 次, 其中取甲盒 $N + 1$ 次, 取乙盒 $N - r$ 次, 且第 $2N - r + 1$ 次是取出甲盒. 若把“取出甲盒”看作“成功”, 则事件 B 就相当于“第 $N + 1$ 次成功之前恰有 $N - r$ 次失败”, 所以, 由例 1.4.10(2)题的结果可得 $P(B) = C_{2N-r}^N \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{N-r}$, 同理可得 $P(C) = C_{2N-r}^N \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{N-r}$, 所以

$$P(A) = P(B) + P(C) = 2C_{2N-r}^N \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{N-r} = C_{2N-r}^N \left(\frac{1}{2}\right)^{2N-r}.$$

(2)设 D = “在第一次用完一盒(不是发现空)时另一盒恰有 r 根”, E = “用完甲盒时乙盒恰有 r 根”, F = “用完乙盒时甲盒恰有 r 根”, 则 $D = E + F$, 且 E 与 F 互不相容, 因此 $P(D) = P(E) + P(F)$, 下面求 $P(E)$ 与 $P(F)$. 如果该数学家用完甲盒的火柴时乙盒恰有 r 根火柴, 那么此人取甲盒和乙盒的次数共有 $2N - r$ 次, 其中取甲盒 N 次, 取乙盒 $N - r$ 次, 且第 $2N - r$ 次是取出甲盒, 因此事件 E 等价于“第 N 次成功以前恰失败了 $N - r$ 次”, 于是由例 1.4.10(2)题的结果可得 $P(E) = C_{2N-r-1}^{N-1} \left(\frac{1}{2}\right)^N \left(\frac{1}{2}\right)^{N-r}$, 同理 $P(F) = C_{2N-r-1}^{N-1} \left(\frac{1}{2}\right)^N \left(\frac{1}{2}\right)^{N-r}$, 所以

$$P(D) = P(E) + P(F) = 2C_{2N-r-1}^{N-1} \left(\frac{1}{2}\right)^N \left(\frac{1}{2}\right)^{N-r} = C_{2N-r-1}^{N-1} \left(\frac{1}{2}\right)^{2N-r-1}.$$

同步练习 1.4

1. 一射手命中 10 环的概率为 0.7, 命中 9 环的概率为 0.3, 求该射手射击三发而命中不少于 29 环的概率.

2. 甲、乙两个篮球运动员, 投篮的命中率分别为 0.7 和 0.8, 如果每人投篮两次, 求:

(1)甲投进两球而乙投进一球的概率;

(2)若投进 1 个球得 2 分, 未投进得 0 分, 求甲、乙两人得分相等的概率.

3. 某车间 10 台同类型的设备, 每台设备的电动机功率为 10 kW, 已知每台设备每小时实际开动 12 min, 它们的使用是相互独立的, 因某种原因, 这天供电部门只能给车间提供 50 kW 的电力, 问该天这 10 台设备能正常运作的概率是多少?

4. 某厂生产的每台仪器, 以概率为 0.7 可以直接出厂, 以概率为 0.3 需进一步调试, 调试后以概率 0.8 可以出厂, 现工厂新生产了 10 台仪器, 设每台仪器的生产是独立的, 求:

(1)仪器全部能出厂的概率;

(2)其中恰有两台不能出厂的概率;

(3)其中至少有两台不能出厂的概率.

5. 一名射手对飞机进行 4 次独立射击, 每次射击命中的概率为 0.3, 一次命中时飞机被击落的概率为 0.6, 至少两次命中时飞机必被击落, 求飞机被击落的概率.

6. 考卷中有一道选择题有 4 个答案, 仅有一个是正确的, 设一个学生知道正确答案或不知道而乱猜是等可能的, 如果这个学生答对了, 求他确实知道正确答案的概率.

7. 已知一批产品 90% 是合格品, 检查产品时, 一个合格品被误认为次品的概率为 0.02, 而一个次品被误认为合格品的概率为 0.05. 求:

- (1) 检查一个产品被认为合格品的概率;
- (2) 被认为合格品的产品确实合格的概率.

8. 为了传递消息, 采用电报系统发出“·”和“-”的信号, 由于电波在传播中会受到随机干扰, 当发出“·”时, 接收为“·”和“-”的概率分别为 0.8 和 0.2, 当发出“-”时, 接收为“·”和“-”的概率分别为 0.1 和 0.9. 现假定发送信号为“·”和“-”的概率为 $\frac{1}{2}$, 又已知收到信号为“·”, 求原来发出的信号是“·”的概率.

9. 某种产品的商标为“MAXAM”, 其中有 2 个字母脱落, 有人捡起随意放回, 求放回后仍为“MAXAM”的概率.

10. 将 A、B、C 三个字母之一输入信道, 输出原字母的概率为 α , 而输出为其他一个字母的概率为 $\frac{1-\alpha}{2}$, 今将字母串 AAAA, BBBB, CCCC 之一输入信道, 输出 AAAA, BBBB, CCCC 的概率分别为 p_1, p_2, p_3 ($p_1 + p_2 + p_3 = 1$), 设信道传输每个字母的工作是独立的, 求:

- (1) 输出为 ABCA 的概率;
- (2) 已知输出为 ABCA, 输入的是 AAAA 的概率.

同步练习 1.4 参考答案

$$1. C_3^0 0.3^0 0.7^3 + C_3^1 0.3^1 0.7^2 = 0.784.$$

$$2. (1) (C_2^0 0.7^2 0.3^0) \cdot (C_2^1 0.8^1 0.2^1) = 0.1568.$$

$$(2) (C_2^0 0.7^2) \cdot (C_2^0 0.8^2) + (C_2^1 0.7^1 0.3^1) (C_2^1 0.8^1 0.2^1) + (C_2^2 0.3^2) \cdot (C_2^2 0.2^2) \\ = 0.4516.$$

$$3. \sum_{k=0}^5 C_{10}^k \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{10-k} = 0.994.$$

$$4. (1) C_{10}^{10} 0.94^{10} 0.06^0 = 0.94^{10}.$$

$$(2) C_{10}^2 0.94^8 0.06^2.$$

$$(3) 1 - (0.94)^{10} - 10 \times 0.94^9 \times 0.06.$$

$$5. 1 - (0.7^4 \times 1 + C_4^1 0.3 \times 0.7^3 \times 0.4) = 0.595.$$

$$6. \frac{\frac{1}{2}}{\frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{4}} = \frac{4}{5}.$$

$$7. (1) 0.9 \times 0.98 + 0.1 \times 0.05 = 0.887.$$

$$(2) \frac{0.9 \times 0.98}{0.887} = 0.994.$$

$$8. 0.889.$$

$$9. \frac{1}{2} \times \frac{2}{10} + \frac{1}{2} \times \frac{2}{10} + \frac{1}{2} \times \frac{4}{10} + 1 \times \frac{1}{10} + 1 \times \frac{1}{10} = \frac{3}{5}.$$

$$10. (1) p_1 \cdot \frac{1}{4} \alpha^2 (1-\alpha)^2 + p_2 \cdot \frac{1}{8} \alpha (1-\alpha)^3 + p_3 \cdot \frac{1}{8} \alpha (1-\alpha)^3 \\ = \frac{1}{8} \alpha (1-\alpha)^2 (3\alpha p_1 - p_1 - \alpha + 1).$$

$$(2) \frac{p_1 \cdot \frac{1}{4} \alpha^2 (1-\alpha)^2}{\frac{1}{8} \alpha (1-\alpha)^2 (3\alpha p_1 - p_1 - \alpha + 1)} = \frac{2\alpha p_1}{3\alpha p_1 - p_1 - \alpha + 1}.$$

疑难解析

1. 在什么情况下, 随机事件 A 的频率可以作为它的概率的近似值?

答 随机事件 A 的频率 $f_n(A)$ 反映事件 A 在多次重复试验中发生的频繁程度. 当 n 增大时, 频率在概率 $P(A)$ 附近摆动, 因此每一个从独立重复试验中测得的频率都可以作为概率 $P(A)$ 的近似值, 而且一般 n 越大, 近似程度越好. 事实上, 当 n 增大时, 频率大量集中于包含 $P(A)$ 的一个小区间, 任选区间中一值作为概率的近似值, 称为统计概率. 在解题时, 当 n 较大时, 可取统计概率为 $P(A) \approx \frac{f_n}{n} = f_n(A)$.

2. 概率是否可以看做频率的极限?

答 这样理解是不恰当的, 因为如上题所述, 当 $n \rightarrow \infty$ 时, $f_n(A)$ 在 $P(A)$ 附近摆动, 与高等数学中极限的“ $\epsilon - N$ ”定义是不同的. 由于概率是随机事件发生的可能性的赋值, 对于任给的 $\epsilon > 0$, 存在偶然的因素, 可能找不到 $N(\epsilon)$, 从而得不到 $|f_n(A) - P(A)| < \epsilon$.

3. 怎样理解古典概型的等可能假设?

答 所讨论的问题是否符合等可能假设, 一般不是通过实际验证, 而往往是根据人们长期形成的“对称性经验”做出的, 例如, 骰子是正六面形的, 当质量均匀分布时, 投掷一次, 每面朝上的可能性都相等; 装在袋中的小球, 颜色可以不同, 只要大小和形状相同, 摸出其中任一个的可能性都相等. 因此, 等可能假设不是人为的, 而是人们根据对事物的认识——对称性特征而确认的.

4. 如何理解对立事件与互不相容事件?

答 如果两个事件 A 与 B 必有一个发生, 且至多有一个发生, 则 A, B 为对立事件, 即 $B = \bar{A}$. 如果两个事件 A 与 B 不能同时发生, 则 A, B 为互不相容事件.

因而, 对立必定互不相容, 互不相容未必对立, 如考试及格与考试不及格既是对立又是互不相容的, 但考 70 分与考 80 分互不相容却不对立.

5. 两个事件 A, B 相互独立与 A, B 互不相容这两个概念有何关系?

答 两个事件 A, B 相互独立, 其实质上是一个事件 B 出现的概率与另一事件 A 是否出现没有关系, 而 A, B 互不相容, 指的是 A 与 B 不同时发生, 从而 B 出现的概率与另一事件 A 是否出现密切相关.

在 $P(A) > 0, P(B) > 0$ 的条件下, 若 A, B 相互独立, 则 $P(AB) = P(A) \cdot P(B) > 0$, 而若 A, B 互不相容, 则 $P(AB) = 0$, 两种概念出现矛盾, 说明在 $P(A) > 0, P(B) > 0$ 的情况下, 相互独立不能互不相容.

6. 在实际应用中, 如何判断两事件的独立性?

答 在实际应用中, 对于事件的独立性常常不是用定义来判断, 而是由试验方式来判断试验的独立性, 由试验的独立性来判断事件的独立性, 或者说根据问题的实质直观上看一事件是否影响另一事件的概率来判断, 例如甲、乙两名运动员在相同条件下投篮, 则“甲投中”与“乙投中”两事件是独立的.

如果对实际问题中的事件还难以判断它们是否独立, 则需要利用统计资料进行分析, 再来

判断是否符合事件独立性的条件.

7. 条件概率 $P(B|A)$ 与无条件概率 $P(B)$ 有什么区别?

答 讨论事件 B 的概率是相对于某组确定的条件 S 而言的, 即 $P(B)$ 是在条件组 S 的实现之下, 事件 B 的概率.

而条件概率 $P(B|A)$ 是在条件组 S 之外, 又附加了一个新条件——事件 A 发生, 求事件 B 发生的概率.

8. 条件概率 $P(B|A)$ 与乘积概率 $P(AB)$ 有什么区别?

答 乘积概率 $P(AB)$ 表示在全部基本事件构成的集合 S 中, 计算事件 AB 的概率, 而条件概率 $P(B|A)$ 表示在原有条件下附加了一个新条件——事件 A 发生后, 在缩小的基本事件构成的集合 S_A 中, 计算 B 发生的概率, 利用古典概率公式, 则有

$$P(B|A) = \frac{AB \text{ 中包含的基本事件数}}{S_A \text{ 中的基本事件数}}.$$

9. 如何使用全概率公式与 Bayes 公式?

答 全概率公式是一个应用广泛的公式, 它把事件 A 的概率(不太好求)分成几个比较容易计算的概率之和, 看似繁琐, 实则简单. 在分析问题的过程中, 把事件 $B_i (i=1, 2, \dots, n)$ 看成 A 发生的原因, A 是结果, 而 $P(B_i)$ 及 $P(A|B_i) (i=1, 2, \dots, n)$ 比较好求, 从而由“原因”求出“结果”.

Bayes 公式有时称为后验概率公式, 它实际上是条件概率, 是在已知结果发生的情况下, 求导致结果的某种原因的可能性大小, 比如求 $P(B_1|A)$, 当 $P(A)$ (常用全概率公式计算)、 $P(B_1)$ 、 $P(A|B_1)$ 比较好求, 就要用 Bayes 公式, 它是由“结果”求“原因”.

10. 什么是小概率事件? 它有什么实际的意义?

答 小概率事件是指一次试验中发生的概率很小的事件, 但可以证明, 一个事件发生的概率不论多小, 只要不断重复试验下去, 事件迟早会出现的概率为 1.

设小概率事件 A 的概率 $P(A) = p > 0$, A_k 表示事件 A 在第 k 次试验中出现, 则 $P(A_k) = p$, $P(\bar{A}_k) = 1 - p$, 于是在前 n 次试验中, A 至少出现一次的概率为

$$P_n = 1 - P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_n) = 1 - (1 - p)^n \rightarrow 1 (n \rightarrow \infty).$$

然而在实践中, 人们总结到“概率很小的事件在一次试验中几乎是不发生的”, 而某一小概率事件在一次试验中竟然发生了, 就有理由怀疑该事件是小概率事件的正确性.

本章常考题型归纳

考点	常考题型	例题
随机事件关系及运算	随机事件间的关系及其运算	例 1.2.1~例 1.2.4
古典概型	古典概型的计算	例 1.1.1~例 1.1.10
利用性质、公式、概型计算概率	利用概率加法公式计算随机事件的概率	例 1.2.5~例 1.2.9
	计算条件概率	例 1.3.1~例 1.3.6
	利用事件的独立性计算概率	例 1.3.7~例 1.3.8
	利用全概率公式和 Bayes 公式计算概率	例 1.4.1~例 1.4.5
	Bernoulli 概型的计算	例 1.4.6~例 1.4.11

- C. “甲种产品滞销” D. “甲种产品滞销或乙种产品畅销”
3. 对于任意两个事件 A, B , 有 $P(A - B)$ 等于().
- A. $P(A) - P(B)$ B. $P(A) - P(B) + P(AB)$
- C. $P(A) - P(AB)$ D. $P(A) + P(\bar{B}) + P(A\bar{B})$
4. 一种零件的加工由两道工序组成, 第一道工序的废品率为 p , 第二道工序的废品率为 q , 则该零件加工的成品率为().
- A. $1 - p - q$ B. $1 - pq$ C. $1 - p - q + pq$ D. $(1 - p) + (1 - q)$
5. 设 $P(A) = 0.8, P(B) = 0.7, P(A|B) = 0.8$, 则下列结论正确的是().
- A. A 与 B 互不相容 B. $A \subset B$
- C. A 与 B 互相独立 D. $P(A + B) = P(A) + P(B)$
6. 袋中有 3 个新球 2 个旧球, 现每次取一个, 无放回地抽取两次, 则第二次取到新球的概率为().
- A. $\frac{3}{5}$ B. $\frac{3}{4}$ C. $\frac{2}{4}$ D. $\frac{3}{10}$
7. 事件 $A(\overline{B+C})$ 的含义是().
- A. A 发生 B. A 发生且 B, C 都不发生
- C. $B + C$ 不发生 D. A 发生, B 和 C 至少有一个不发生
8. 五个考签中有一个难签, 甲、乙、丙三个考生依次从中抽取出一张考签, 设他们抽到难签的概率分别为 p_1, p_2, p_3 , 则().
- A. $p_1 < p_2 < p_3$ B. $p_1 = p_2 = p_3$ C. $p_1 > p_2 > p_3$ D. 不能排大小
9. 同时掷 3 枚均匀硬币, 恰有两枚正面朝上的概率为().
- A. 0.5 B. 0.25 C. 0.125 D. 0.375
10. 设 $0 < P(A) < 1, P(B) > 0$, 且 $P(B|A) = P(B|\bar{A})$, 则有().
- A. $P(A|B) = P(\bar{A}|B)$ B. $P(A|B) \neq P(\bar{A}|B)$
- C. $P(AB) = P(A)P(B)$ D. $P(AB) \neq P(A)P(B)$

二、填空题

1. 已知 $P(A) = 0.4, P(B) = 0.3$.
- (1) 当 A, B 互不相容时, $P(A + B) = \underline{\hspace{2cm}}, P(AB) = \underline{\hspace{2cm}}$.
- (2) 当 A, B 相互独立时, $P(A + B) = \underline{\hspace{2cm}}, P(AB) = \underline{\hspace{2cm}}$.
- (3) 当 $B \subset A$ 时, $P(A + B) = \underline{\hspace{2cm}}, P(AB) = \underline{\hspace{2cm}}$.
2. 已知 $P(A) = 0.5, P(B) = 0.6, P(B|A) = 0.8$, 则 $P(A + B) = \underline{\hspace{2cm}}$.
3. 设在全部产品中 2% 是废品, 而合格品有 85% 是一级品, 则任抽出一件产品是一级品的概率为 $\underline{\hspace{2cm}}$.
4. 设在 1 000 件产品中有 200 件是不合格品, 依次做无放回抽取两件产品, 则第二次抽取到不合格品的概率是 $\underline{\hspace{2cm}}$.
5. 某人投篮四次, 已知四次中至少投中一次的概率为 0.998 4, 则他投篮的命中率为 $\underline{\hspace{2cm}}$.

三、计算题

1. 已知 10 个晶体管中有 7 个正品 3 个次品, 每次任意抽取一个来测试, 测试后不再放回

去,直至把3个次品都找到为止,求需要测试7次的概率.

2. 为了了解儿童对电视商品广告的理解能力,电视台做了一次随机抽样调查,在不同的年龄组共选取1 000名儿童,情况如下:

程度 \ 年龄(岁)	4~6	7~9	10~12
理解(名)	81	185	231
不理解(名)	410	73	20

现随机抽出一名儿童发现他不理解广告,求他是7~9岁儿童的概率.

3. 一批产品共100件,其中有4件次品,现从中任取3件来检验,如发现有次品,则认为这批产品不合格,但检验时,一件正品被误判为次品的概率为0.05,而一件次品被误判为正品的概率为0.01,求这批产品是合格品的概率.

4. 设有白球与黑球各4只,从中任取4只放入甲盒,余下的4只放入乙盒,然后分别在两盒中各任取1球,颜色正好相同,试问放入甲盒的4只球中有几只白球的概率最大?并求此概率值.

5. 甲、乙两名篮球运动员,投篮命中率分别为0.7和0.6,每人投篮3次,求:

(1)二人进球数相等的概率;

(2)甲比乙进球数多的概率.

四、证明题

1. 若事件 A 、 B 互不相容,求证: $P(A\bar{B})=P(A)$.

2. 已知事件 A 、 B 、 C 中任意一个事件发生时事件 D 发生,求证:

$$P(D) \geq P(A) + P(B) + P(C) - 2.$$

3. 若事件 A 、 C 独立, B 、 C 独立,且 $B \subset A$,求证: $A - B$ 与 C 独立.

检测题(一)参考答案

一、选择题

1. A 2. D 3. C 4. C 5. C
6. A 7. C 8. B 9. D 10. C

二、填空题

1. (1)0.7 0 (2)0.58 0.12 (3)0.4 0.3
2. 0.7 3. 0.83 4. 0.2 5. 0.8

三、计算题

$$1. \frac{C_6^2 C_4^4 A_7^4 3!}{A_{10}^7} = 0.125.$$

$$2. \frac{73}{0.503} = 0.5625.$$

$$3. \frac{1}{C_{100}^3} (0.95^3 \times C_4^0 C_{96}^3 + 0.01 \times 0.95^2 \times C_4^1 C_{96}^2 + 0.01^2 \times 0.95 \times C_4^2 C_{96}^1 + 0.01^3 \times C_4^3 C_{96}^0) =$$

0.7586.

4. 有2只白球的概率最大,其值为 $\frac{3}{5}$.

5. (1) 0.321. (2) 0.436.

四、证明

$$1. P(A) = P(AB + A\bar{B}) = P(AB) + P(A\bar{B}) = P(A\bar{B}).$$

$$2. A \subset D, B \subset D, C \subset D, P(D) \geq P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \geq P(A) + P(B) + P(C) - 2.$$

3. $B \subset A$ 则 $AB = B$.

$$P((A - B)C) = P(AC - BC) = P(AC) - P(ABC) = P(AC) - P(BC) = P(A)P(C) - P(B)P(C) = (P(A) - P(B))P(C) = P(A - B)P(C).$$

检测题(二)

一、选择题

1. 设随机事件 A, B, C 两两互不相容, $P(A) = 0.2, P(B) = 0.3, P(C) = 0.4$, 则 $P((A + B) - C)$ 等于().

- A. 0.5 B. 0.1 C. 0.44 D. 0.3

2. 设当事件 A 与 B 同时发生时, 事件 C 必发生, 则下列各式正确的是().

- A. $P(C) \leq P(A) + P(B) - 1$ B. $P(C) \geq P(A) + P(B) - 1$
C. $P(C) = P(AB)$ D. $P(C) = P(A + B)$

3. 设 $0 < P(A) < 1, 0 < P(B) < 1, P(A|B) + P(\bar{A}|\bar{B}) = 1$, 则下列各式正确的是().

- A. 事件 A 和 B 互不相容 B. 事件 A 和 B 互相对立
C. 事件 A 和 B 互不独立 D. 事件 A 和 B 相互独立

4. 设 A, B 为任意两个事件, 且 $A \subset B, P(B) > 0$, 则下列选项必须成立的是().

- A. $P(A) < P(A|B)$ B. $P(A) \leq P(A|B)$
C. $P(A) > P(A|B)$ D. $P(A) \geq P(A|B)$

5. 每次试验成功率为 $p (0 < p < 1)$, 进行重复实验, 直到第 10 次试验才取得 4 次成功的概率为().

- A. $C_{10}^4 p^4 (1-p)^6$ B. $C_9^3 p^4 (1-p)^6$
C. $C_9^4 p^4 (1-p)^5$ D. $C_9^3 p^3 (1-p)^6$

6. 设 $P(A) = a, P(B) = b, P(A + B) = c$, 则 $P(A\bar{B})$ 等于().

- A. $a(1-b)$ B. $a-b$ C. $c-b$ D. $a(1-c)$

7. 10 张奖券中含有 3 张中奖的奖券, 每人购买一张, 则前 3 名购买者中恰有 1 人中奖的概率为().

- A. $C_{10}^3 \cdot (0.7)^2 \cdot (0.3)^1$ B. 0.3 C. $\frac{7}{40}$ D. $\frac{21}{40}$

8. 设事件 A, B, C 相互独立, 且 $0 < P(C) < 1$, 则在下列四对事件中不相互独立的是().

- A. $\overline{A+B}$ 与 $\bar{C}$ B. $\overline{AC}$ 与 $\bar{C}$
C. $\overline{A-B}$ 与 $\bar{C}$ D. $\overline{AB}$ 与 $\bar{C}$

9. 已知 $0 < P(B) < 1$, 且 $P((A_1 + A_2)|B) = P(A_1|B) + P(A_2|B)$, 则下列成立的是().

A. $P((A_1 + A_2)|\bar{B}) = P(A_1|\bar{B}) + P(A_2|\bar{B})$

$$B. P(A_1B + A_2B) = P(A_1B) + P(A_2B)$$

$$C. P(A_1 + A_2) = P(A_1|B) + P(A_2|B)$$

$$D. P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2)$$

10. 设 A, B, C 是两两相互独立且不可能同时发生的事件, 且 $P(A) = P(B) = P(C) = p$, 则 p 的最大值为().

$$A. \frac{1}{2}$$

$$B. 1$$

$$C. \frac{1}{3}$$

$$D. \frac{1}{4}$$

二、填空题

1. 已知 $P(A) = P(B) = P(C) = \frac{1}{4}$, $P(AB) = 0$, $P(AC) = P(BC) = \frac{1}{6}$, 则事件 A, B, C 全不发生的概率为_____.

2. 某射手在三次射击中至少命中一次的概率为 0.875, 则这名射手在三次射击中至多命中一次的概率为_____.

3. 若事件 A, B 满足 $P(AB) = P(\overline{AB})$, 且 $P(A) = p$, 则 $P(B) =$ _____.

4. 某种集成电路使用到 3 000 h 还能正常工作的概率为 0.95, 使用到 4 000 h 还能正常工作的概率为 0.8, 则已经工作了 3 000 h 的集成电路, 还能再正常工作 1 000 h 的概率为_____.

5. 一批产品中一、二、三等品各占 60%、30%、10%, 从中任意取出一件, 结果不是三等品, 则取到的是一等品的概率为_____.

三、计算题

1. 一次掷 10 颗骰子, 已知至少出现一个一点, 求至少出现两个一点的概率是多少?

2. 50 个铆钉中有 3 个强度太弱, 假若这 3 个铆钉都装在同一个部件上, 则这个部件的强度就太弱, 现有 10 个部件, 每个部件装 3 个铆钉, 若从 50 个铆钉中随机取用, 问恰有 1 个部件强度太弱的概率是多少?

3. 一学生接连参加同一课程的两次考试, 第一次考试及格的概率为 p , 若第一次及格则第二次及格的概率也为 p , 若第一次不及格则第二次及格的概率为 $\frac{p}{2}$.

(1) 若至少有一次及格则他能取得某种资格, 求他取得该资格的概率;

(2) 若已知他第二次已经及格, 求他第一次及格的概率.

4. 某保险公司认为人可以分为两类, 第一类是容易出事故的, 另一类则比较谨慎. 在任何给定的一年期间, 易出事故的人发生一起事故的概率为 0.4, 而对比较谨慎的人而言, 这个概率为 0.2, 若假定第一类人占人口的 30%, 设已知某被保险人在第一年已出过一次事故, 问他在保险有效的第二年又出一次事故的概率是多少?

5. 假设每个飞机引擎在练习飞行中出故障的概率为 p , 有多引擎飞机的各引擎是否出故障是独立的, 如果飞机正常运行至少需要 50% 的引擎正常工作, 问当 p 满足什么条件时, 4 引擎飞机比 2 引擎飞机更安全?

四、证明题

1. 求证: 对任意事件 A, B 有 $P(A+B)P(AB) \leq P(A)P(B)$.

2. 求证: 对任意事件 A, B , $P(B)$ 介于 $P(B|A)$ 和 $P(B|\bar{A})$ 之间.

检测题(二)参考答案

一、选择题

1. A 2. B 3. D 4. B 5. B
6. C 7. C 8. B 9. B 10. A

二、填空题

1. $\frac{3}{8}$ 2. $\frac{1}{2}$ 3. $1-p$ 4. $\frac{16}{19}$ 5. $\frac{2}{3}$

三、计算题

$$1. 1 - \frac{C_{10}^1 \cdot 5^9}{6^{10}} = 0.6148.$$

$$2. \frac{10C_{47}^{27} \frac{27!}{3!}}{C_{50}^{30} (3!)^{10}} = \frac{1}{1960}. \quad 3. (1) \frac{3}{2}p - \frac{1}{2}p^2. (2) \frac{2p}{1+p}.$$

4. 设 A_i = “被保险人在第 i 年出事故” ($i=1, 2$),
 B = “被保险人是第一类人”.

$$\begin{aligned} P(A_2|A_1) &= \frac{P(A_1A_2)}{P(A_1)} = \frac{P((A_1B + A_1\bar{B})A_2)}{P(A_1)} = \frac{P(A_1BA_2 + A_1\bar{B}A_2)}{P(A_1)} \\ &= \frac{P(A_1BA_2)}{P(A_1)} + \frac{P(A_1\bar{B}A_2)}{P(A_1)} = \frac{P(A_1B)}{P(A_1)}P(A_2|A_1B) + \frac{P(A_1\bar{B})}{P(A_1)}P(A_2|A_1\bar{B}) \\ &= P(B|A_1)P(A_2|A_1B) + P(\bar{B}|A_1)P(A_2|A_1\bar{B}) = \frac{6}{13} \times 0.4 + \frac{7}{13} \times 0.2 \\ &= 0.2923. \end{aligned}$$

$$5. C_4^2 p^2 (1-p)^2 + C_4^3 p^3 (1-p) + C_4^4 p^4 > C_2^1 p (1-p) + C_2^2 p^2,$$

即 $p > \frac{2}{3}.$

四、证明题

1. (1) 在 $A \subset B, B \subset A, AB = \phi, P(A)=0, P(B)=0$ 的特殊情况下, 不等式显然成立.

$$(2) P(\bar{A}B)P(AB) \leq P(A)P(\bar{A}B),$$

$$\begin{aligned} P(A)P(AB) + P(\bar{A}B)P(AB) &\leq P(A)P(AB) + P(A)P(\bar{A}B), \\ (P(A) + P(\bar{A}B))P(AB) &\leq P(A)(P(AB) + P(\bar{A}B)). \end{aligned}$$

由 $A + B = A + \bar{A}B, B = AB + \bar{A}B, A(\bar{A}B) = \emptyset, (AB)(\bar{A}B) = \emptyset$, 所以

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

2. (1) 当 $P(AB) \leq P(A)P(B)$ 时,

$$P(B) - P(AB) \geq P(B) - P(A)P(B), \text{ 即 } P(B)P(\bar{A}) \leq P(\bar{A}B),$$

故 $\frac{P(AB)}{P(A)} \leq P(B) \leq \frac{P(\bar{A}B)}{P(\bar{A})}$, 即 $P(B|A) \leq P(B) \leq P(B|\bar{A})$.

- (2) 当 $P(AB) \geq P(A)P(B)$ 时,

类似可证得 $P(B|\bar{A}) \leq P(B) \leq P(B|A)$.

第2章 随机变量及其分布

2.1 离散型随机变量

2.1.1 知识要点

1. 随机变量

①定义:如果在一定条件 S 下的随机试验的每一个可能结果 ω 都唯一地对应一个实数值,则称此实值变量为一个随机变量,记为 $X(\omega)$ 或简记为 X .

②常见随机变量类型:离散型随机变量,连续型随机变量.

2. 离散型随机变量

定义:如果随机变量 X 全部可以取到的值能够一一列举出来,则称 X 为离散型随机变量.

3. 概率分布

①定义:设离散型随机变量 X 可能取到的值为 $x_1, x_2, \dots, x_k, \dots$, 且取各个值 x_k 的概率为 $P(X=x_k)$. 如果记 $p_k = P(X=x_k) (k=1, 2, \dots)$, 则称此式为 X 的概率分布,或称为分布律.

分布律还可采用表格的形式表示:

X	x_1	x_2	x_3	$\dots$	x_k	$\dots$
P	p_1	p_2	p_3	$\dots$	p_k	$\dots$

这个表也称为 X 的概率分布表.

②性质: $p_k \geq 0 (k=1, 2, \dots)$;

$$\sum_k p_k = 1.$$

4. 常见的概率分布

①两点分布(0—1 分布)

X	0	1
P	$1-p$	p

($0 < p < 1$).

记为 $X \sim B(1, p)$.

②二项分布(Bernoulli 分布)

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

记为 $X \sim B(n, p)$.

③Poisson 分布

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

记为 $X \sim P(\lambda)$.

④超几何分布

$$P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \quad (k=0, 1, 2, \dots, L; L = \min\{M, n\}).$$

5. 两个定理

(1)二项分布与 Poisson 分布的关系

定理 1 (Poisson 定理) 设随机变量 $X_n (n=1, 2, \dots)$ 服从二项分布, 其分布律为 $P(X_n = k) = C_n^k p^k (1-p)^{n-k} (k=0, 1, 2, \dots, n)$, 且 p_n 是仅与 n 有关的数, $np_n = \lambda$ 是常数 ($n=1, 2, \dots$) 则有

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

注 Poisson 定理表明, 当 n 很大, p 很小时, 有近似公式

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \quad (\lambda = np).$$

在实际计算中, 当 $n \geq 10, p \leq 0.1$ 时, 就可用上述近似公式来进行二项分布的近似计算, 以达到简化计算的目的.

(2)二项分布与超几何分布的关系

定理 2 设 n, k 均为不变的正整数, 如果 $\lim_{N \rightarrow \infty} \frac{M}{N} = p$ (N, M 均为正整数) 则有

$$\lim_{N \rightarrow \infty} \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = C_n^k \cdot p^k \cdot (1-p)^{n-k}.$$

注 此定理表明, 当 N 很大时, 有近似公式 $\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k}$, 即可用二项分布作为超几何分布的近似计算.

2.1.2 典型题型解析

1. 有关离散型随机变量概率分布的性质的题

例 2.1.1 某人求出下面列出的概率分布

$$(1) \begin{array}{c|cccc} X & 1 & 2 & 3 & 4 \\ \hline P & 0.1 & 0.4 & 0.3 & 0.2 \end{array};$$

$$(2) \begin{array}{c|ccccc} X & 1 & 2 & 3 & 4 & \dots \\ \hline P & \frac{1}{2} & \frac{1}{2 \cdot 4} & \frac{1}{2 \cdot 4^2} & \frac{1}{2 \cdot 4^3} & \dots \end{array}.$$

问是否为离散型随机变量的概率分布?

解 (1) X 有有限多个值, $0 \leq p_k \leq 1$, 且 $\sum p_k = 0.1 + 0.4 + 0.3 + 0.2 = 1$, 所以是离散型随机变量的概率分布.

(2) X 有可列无限多个值, $0 \leq p_k = \frac{1}{2 \times 4^{k-1}} \leq 1$, 但是 $\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} \frac{1}{2 \cdot 4^{k-1}} = \frac{2}{3} \neq 1$, 所以不是离散型随机变量的概率分布.

例 2.1.2 问 c 为何值时, $P(X=k) = c \cdot \frac{\lambda^k}{k!} (k=1, 2, \dots; \lambda > 0)$ 表示一个离散型随机变量的分布律?

解 $\sum_{k=1}^{\infty} P(X=k) = \sum_{k=1}^{\infty} c \cdot \frac{\lambda^k}{k!} = c \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - 1 \right) = c(e^\lambda - 1)$, 所以 $c = (e^\lambda - 1)^{-1}$; 又因为 $\lambda > 0$, 因此 $c = (e^\lambda - 1)^{-1} > 0$, 所以对于一切的 $k=1, 2, \dots$ 均有 $P(X=k) > 0$. 综上所述可知, 当 $c = (e^\lambda - 1)^{-1}$ 时, $P(X=k) = c \cdot \frac{\lambda^k}{k!}$ 表示一个离散型随机变量的分布律.

注 要确定或判别某等式 $P(X=k) = p_k (k=1, 2, \dots)$ 是否表示某一离散型随机变量的概率分布, 其充分必要条件是:

(1) $p_k \geq 0 (k=1, 2, \dots)$;

(2) $\sum_k p_k = 1$.

2. 求离散型随机变量的概率分布

例 2.1.3 一袋中有 4 个球, 分别标有数字 1, 2, 3, 4, 从中任取 2 个球, 求最小号码的概率分布.

解 设 X 表示取出的 2 个球中的最小号码, 则 X 的取值为 1, 2, 3. 基本事件总数为 $C_4^2 = 6$. “ $X=1$ ”表示“取出的 2 个球中最小号码为 1”, 包含 3 个基本事件, 即(1, 2), (1, 3), (1, 4); “ $X=2$ ”表示“取出的 2 个球中最小号码为 2”, 包含 2 个基本事件, 即(2, 3), (2, 4); “ $X=3$ ”表示“取出的 2 个球中最小号码为 3”, 包含 1 个基本事件, 即(3, 4). 所以 X 的概率分布为

X	1	2	3
P	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

例 2.1.4 把 3 个乒乓球放入 4 个盒子中, 求盒子中盛乒乓球数最多的概率分布.

解 设 X 表示盒子中盛乒乓球的最多个数, 则 X 的取值为 1, 2, 3. 基本事件总数为 4^3 . 事件“ $X=1$ ”表示“4 个盒子中恰有 3 个盒子中各有一个乒乓球”, 包含 A_4^3 个基本事件; 事件“ $X=2$ ”表示“4 个盒子中恰有一个盒子中有 2 个乒乓球, 另外 3 个盒子中, 有 1 个盒子里有 1 个乒乓球”, 包含 $C_4^1 \cdot C_3^2 \cdot C_3^1$ 个基本事件; 事件“ $X=3$ ”表示“4 个盒子中恰有 1 个盒子中有 3 个乒乓球”, 包含 $C_4^1 \cdot C_3^3$ 个基本事件, 所以 X 的概率分布为

X	1	2	3
P	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{1}{16}$

例 2.1.5 在 n 重 Bernoulli 试验中, 每次试验 A 发生的概率为 p . 设 X 表示 A 发生 k 次所需进行的试验次数, 求 X 的概率分布.

解 X 的取值为 $k, k+1, \dots$. 事件“ $X=m$ ”表示在前 $m-1$ 次试验中 A 发生了 $k-1$ 次,

第 m 次试验 A 发生了, 所以 X 的概率分布为

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

这时称 X 服从以 k, p 为参数的帕斯卡分布, 特别地, 当 $k=1$ 时, 称 X 服从以 p 为参数的几何分布.

注 求一个离散型随机变量的概率分布的思路是: 首先确定随机变量的取值, 其次求出相应的概率, 最后按规范形式写出概率分布.

随机变量的取值一般可由具体问题的题意写出, 关键是求 X 取每个值的概率. 为此, 必须要弄清“ $X=x_k$ ”的实际含义, 用具体的随机事件 A_k 表示. 因此求 $P(X=x_k)$ 的问题就是求随机事件 A_k 的概率 $P(A_k)$.

3. 利用离散型随机变量的分布计算概率

例 2.1.6 某人掷硬币 10 次, 写出国徽向上次数的分布律, 并求国徽向上次数不小于 3 的概率.

解 如果在单次试验中, 只有两个可能的结果 A 及 $\bar{A}$, A 发生的概率为 p , 则在 n 次独立试验中事件 A 发生的概率 $p(A \text{ 恰好发生 } k \text{ 次}) = C_n^k p^k (1-p)^{n-k} (k=0, 1, 2, \dots, n)$. 如果以 X 表示 n 重独立试验中事件 A 发生的次数, 则 X 为离散型随机变量, 且服从参数为 n, p 的二项分布, 即 $X \sim B(n, p)$. 所以在本题中设 X 表示国徽向上的次数, 则 $X \sim B\left(10, \frac{1}{2}\right)$, 所以

$$\begin{aligned} P(X \geq 3) &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - C_{10}^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10} - C_{10}^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^9 - C_{10}^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^8 = 0.945. \end{aligned}$$

例 2.1.7 超市订购 1 000 瓶酒, 在运输中瓶子被打破的概率为 0.004. 求超市收到的酒中:

- (1) 恰有 2 瓶被打破的概率;
- (2) 至少有一瓶被打破的概率.

解 设 X 表示被打破的瓶子数, 则 $X \sim B(1\,000, 0.004)$. 由于 $n=1\,000, p=0.004, \lambda=np=4$, 所以可用 Poisson 分布近似计算.

$$(1) P(X=2) = C_{1\,000}^2 \cdot (0.004)^2 \cdot (1-0.004)^{1\,000-2} \approx \frac{4^2 \cdot e^{-4}}{2!} \approx 0.1465;$$

$$\begin{aligned} (2) P(X \geq 1) &= 1 - P(X=0) = 1 - C_{1\,000}^0 (0.004)^0 (1-0.004)^{1\,000} \\ &\approx 1 - \frac{4^0 \cdot e^{-4}}{0!} \approx 0.9817. \end{aligned}$$

例 2.1.8 设 1 小时内进入某图书馆的读者人数服从 Poisson 分布, 已知 1 小时内无人进入图书馆的概率为 0.01. 求 1 小时内至少有 2 个读者进入图书馆的概率.

解 设 X 表示 1 小时内进入图书馆的读者人数, 则 $X \sim p(\lambda)$, 即 $p(X=k) = \frac{\lambda^k e^{-\lambda}}{k!} (k=0, 1, \dots)$. 已知 $P(X=0) = e^{-\lambda} = 0.01$, 因此 $\lambda = 2 \ln 10$. 所以所求概率为

$$P(X \geq 2) = 1 - P(X=0) - P(X=1) = 1 - 0.01(1 + 2 \ln 10) = 0.944.$$

例 2.1.9 由一家商店过去销售记录知, 某种商品每月的销售量可用参数 $\lambda=9$ 的 Poisson 分布描述, 为了有 95% 以上的把握保证不脱销, 问商店在月初进货时要进这种商品多少件?

解 设 X 表示商店每月销售某种商品的件数, 由题意知 $X \sim P(9)$. 设月初进货 a 件就可

以保证以 95% 以上的概率不脱销, 因此 $P(X \leq a) \geq 0.95$, 即 $\sum_{k=0}^a \frac{9^k}{k!} e^{-9} \geq 0.95$, $a = 18$. 所以这家商店只要在月初进该商品 18 件, 就有 95% 以上的把握保证这种商品在本月不脱销.

例 2.1.10 为了保证设备正常工作, 需配备适量的维修工人. 现有同类型设备 300 台, 各台工作是相互独立的, 发生故障的概率都是 0.01. 如果 1 台设备的故障可由一人来处理, 问至少需要配备多少工人, 才能保证当设备发生故障而不能及时维修的概率小于 0.01?

解 设 X 表示同时发生故障的设备数, 由题意知 $X \sim B(300, 0.01)$. 若配备 n 名维修工人, 则根据题目要求有 $P(X > n) < 0.01$, 即

$$\begin{aligned} P(X > n) &= 1 - P(X \leq n) = 1 - \sum_{k=0}^n C_{300}^k (0.01)^k \times (0.99)^{300-k} \\ &\approx 1 - \sum_{k=0}^n \frac{3^k e^{-3}}{k!} < 0.01, \\ \sum_{k=0}^n \frac{3^k e^{-3}}{k!} &> 0.99, \end{aligned}$$

经计算可得 $n = 8$. 所以至少应配备 8 名维修工人.

例 2.1.11 设随机变量 X 的概率分布为 $P(X = k) = \frac{k}{15}$, $k = 1, 2, 3, 4, 5$.

求 (1) $P\left(\frac{1}{2} < X < \frac{5}{2}\right)$; (2) $P(1 \leq X \leq 3)$; (3) $P(X > 3)$.

解 (1) $P\left(\frac{1}{2} < X < \frac{5}{2}\right) = P(X = 1) + P(X = 2) = \frac{1}{15} + \frac{2}{15} = \frac{1}{5}$;

(2) $P(1 \leq X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} = \frac{2}{5}$;

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

例 2.1.12 设随机变量 X 服从参数为 n, p 的二项分布, 问 k 为何值时, $p(X = k)$ 最大.

解 因为对 $0 < p < 1$, 有

$$\frac{C_n^k p^k (1-p)^{n-k}}{C_n^{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)} = 1 + \frac{(n+1)p - k}{k(1-p)}.$$

当 $k < (n+1)p$ 时 $C_n^k p^k (1-p)^{n-k} > C_n^{k-1} p^{k-1} (1-p)^{n-k+1}$;

当 $k = (n+1)p$ 时 $C_n^k p^k (1-p)^{n-k} = C_n^{k-1} p^{k-1} (1-p)^{n-k+1}$;

当 $k > (n+1)p$ 时 $C_n^k p^k (1-p)^{n-k} < C_n^{k-1} p^{k-1} (1-p)^{n-k+1}$.

所以当 $k = \begin{cases} (n+1)p - 1 \text{ 或 } (n+1)p, & (n+1)p \text{ 是整数} \\ [(n+1)p], & (n+1)p \text{ 不是整数} \end{cases}$ 时, $P(X = k)$ 最大.

同步练习 2.1

1. 已知离散型随机变量 X 的概率分布为 $P(X = k) = a \left(\frac{2}{3}\right)^k$ ($k = 1, 2, \dots$). 求 a 的值.

2. 设 X 为离散型随机变量, 其概率分布为

X	-1	0	1
P	$\frac{1}{2}$	$1-2q$	q^2

求 q 的值.

3. 设离散型随机变量 X 的概率分布为 $\begin{array}{c|ccc} X & -1 & 1 & 2 \\ \hline P & 0.2 & 0.5 & 0.3 \end{array}$, 求: (1) $P\left(X > \frac{1}{2}\right)$; (2) $P\left(-1 \leq X < \frac{3}{2}\right)$.
4. 设随机变量 $X \sim P(\lambda)$, 且 $P(X=1) = P(X=2)$. 求 $P(X > 1)$.
5. 一批灯泡共有 40 只, 其中 3 只坏的 37 只好的. 现从中任意抽取 4 只进行检验, 求坏灯泡只数的概率分布.
6. 同时掷两颗骰子, 求两颗骰子与点数之和的概率分布.
7. 一批产品共 10 件, 其中 7 件正品, 3 件次品, 每次从这批产品中任取一次, 在下述三种情况下, 分别求直至取得正品为止所需次数 X 的概率分布:
- (1) 每次取出的产品不再放回;
 - (2) 每次取出的产品仍放回;
 - (3) 每次取出一件产品后, 总是另取一件正品放回这批产品中.
8. 有一繁忙的汽车站, 每天有大量的汽车通过, 设每辆汽车在一天的某段时间内出事故的概率为 0.000 1, 在某天的该段时间内有 1 000 辆汽车通过. 问出事故次数不小于 2 的概率是多少?
9. 某车间有 20 部同型号机床, 每部机床开动的概率为 0.8. 若假定每部机床是否开动彼此独立, 每部机床开动时所耗的电能为 15 个单位, 求这个车间消耗电能不少于 270 个单位的概率.
10. 在保险公司里有 2 500 名同一年龄和同一社会阶层的人参加了保险, 在 1 年中每个人死亡的概率为 0.002, 每个参加保险的人在 1 月 1 日需交 12 元保险费, 而在死亡时家属可从保险公司领取 2 000 元赔偿金. 求:
- (1) 保险公司亏本的概率;
 - (2) 保险公司获利不小于 10 000 元的概率.
11. 一电话交换台有 300 台分机, 假设每台分机要外线的概率为 3%, 问需要设几条外线, 才能使每台分机呼叫外线, 以 70% 的概率得到满足?
12. 设离散型随机变量 X 服从参数为 λ 的 Poisson 分布, 问当 k 取何值时, $P(X=k)$ 最大.

同步练习 2.1 参考答案

1. $a = \frac{1}{2}$.
2. $q = 1 - \frac{\sqrt{2}}{2}$.
3. $p\left(X > \frac{1}{2}\right) = 0.8$, $p\left(-1 \leq X < \frac{3}{2}\right) = 0.7$.
4. $P(X > 1) = 1 - P(X \leq 1) = 1 - 3e^{-2}$.
5. $P(X=k) = \frac{C_3^k C_{37}^{4-k}}{C_{40}^4} (k=0, 1, 2, 3)$.

6.	X	2	3	4	5	6	7	8	9	10	11	12
	P	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

7. (1)	X	1	2	3	4
	P	$\frac{7}{10}$	$\frac{7}{30}$	$\frac{7}{120}$	$\frac{1}{120}$

$$(2) P(X=k) = \frac{7}{10} \left(\frac{3}{10}\right)^{k-1} \quad (k=1, 2, \dots).$$

(3)	X	1	2	3	4
	P	$\frac{7}{10}$	$\frac{24}{100}$	$\frac{54}{1\,000}$	$\frac{6}{1\,000}$

$$8. 1 - e^{-0.1} - 0.1e^{-0.1} = 0.0047.$$

$$9. C_{20}^{18} 0.8^{18} 0.2^2 + C_{20}^{19} 0.8^{19} 0.2 + 0.8^{20} = 0.206.$$

10. 设 X 表示一年中死亡的人数.

$$(1) P(X > 15) = 1 - P(X \leq 15) = 0.000\,069.$$

$$(2) P(X \leq 10) = 0.986\,305.$$

$$11. P(X \leq m) = \sum_{k=0}^m C_{300}^k (0.03)^k (0.97)^{300-k} \approx \sum_{k=0}^m \frac{9^k}{k!} e^{-9} \geq 0.7 \text{ 所以 } m = 10.$$

$$12. k = \begin{cases} \lambda - 1 \text{ 或 } \lambda, & \lambda \text{ 为整数,} \\ [\lambda], & \lambda \text{ 不为整数.} \end{cases}$$

2.2 连续型随机变量

2.2.1 知识要点

1. 定义 1

如果随机变量 X 全部可能取到的值不能一一列举出来, 则称 X 为非离散型随机变量.

2. 定义 2

对于随机变量 X , 如果存在定义在区间 $-\infty < x < +\infty$ 上的非负可积函数 $p(x)$, 使对任意实数 $a, b (a < b)$ 均有

$$P(a < X < b) = \int_a^b p(x) dx,$$

则称 X 为连续型随机变量, 称 $p(x)$ 为 X 的概率密度函数, 也称为概率密度或密度函数, 或简称为密度.

注 非离散型随机变量不都是连续型随机变量, 但在实际问题中遇到的非离散型随机变量大多是连续型的.

3. 概率密度函数的性质

$$\textcircled{1} p(x) \geq 0;$$

$$\textcircled{2} \int_{-\infty}^{+\infty} p(x) dx = 1 ;$$

$$\textcircled{3} P(X=a)=0(a \text{ 为任意实数}),$$

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

4. 常见的概率密度函数

(1) 均匀分布

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

X 服从区间 $[a, b]$ 上的均匀分布, 记为 $X \sim V[a, b]$.

(2) 指数分布

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

X 服从参数为 λ 的指数分布, 记为 $X \sim \Gamma(1, \lambda)$.

(3) 正态分布(Gauss 分布)

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad (-\infty < x < +\infty, \sigma > 0).$$

X 服从参数为 σ, μ 的正态分布, 记为 $X \sim N(\mu, \sigma^2)$.

注 ① 当 $\sigma=1, \mu=0$ 时, 称 X 服从标准正态分布, 记为 $X \sim N(0, 1)$, 其概率密度函数为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (-\infty < x < +\infty).$$

② $p(x)$ 的图形关于直线 $x=\mu$ 对称.

③ $p(x) = p(\mu) = \frac{1}{\sigma \sqrt{2\pi}}$ 为 $p(x)$ 的最大值.

④ 当 $x = \mu \pm \sigma$ 时, 曲线 $y = p(x)$ 有拐点 $(\mu \pm \sigma, \frac{1}{\sigma \sqrt{2\pi e}})$, 且以 Ox 轴为渐近线. 当 σ 大

时, 曲线平缓; 当 σ 小时, 曲线陡峭.

⑤ 如果固定 σ 的值, 改变 μ 的值, 则 $p(x)$ 的图形沿 Ox 轴平行移动, 而不改变其形状.

(4) Γ -分布

$$p(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0, \\ 0, & x \leq 0. \end{cases} \text{ 其中 } \alpha, \beta \text{ 为常数, } \alpha > 0, \beta > 0.$$

X 服从参数为 α, β 的 Γ 分布, 记为 $X \sim \Gamma(\alpha, \beta)$.

注 ① $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx (\alpha > 0)$,

$$\Gamma(1) = 1, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \Gamma(\alpha+1) = \alpha \Gamma(\alpha),$$

$\Gamma(n+1) = n!$ (n 为正整数).

② 若 $\alpha=1$, 则 X 服从参数为 β 的指数分布.

③若 $\alpha = \frac{n}{2}, \beta = \frac{1}{2}$ 则 X 服从自由度为 n 的 χ^2 分布 (n 为正整数), 记为 $X \sim \chi^2(n)$ 其概率密度函数为

$$k_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

2.2.2 典型题型解析

1. 连续型随机变量概率密度的性质

例 2.2.1 下列函数是否为连续型随机变量的概率密度函数:

$$\begin{aligned} (1) p_1(x) &= \begin{cases} \frac{1}{1+x^2}, & x > 0, \\ 0, & x \leq 0; \end{cases} & (2) p_2(x) &= \begin{cases} \frac{1}{2} \cos x, & 0 < x < \pi, \\ 0, & \text{其他}; \end{cases} \\ (3) p_3(x) &= \begin{cases} e^{-(x-a)}, & x > a, \\ 0, & \text{其他}. \end{cases} \end{aligned}$$

解 连续型随机变量的概率密度函数必须是非负函数, 且 $\int_{-\infty}^{+\infty} p(x) dx = 1$.

(1) $p_1(x) \geq 0 (-\infty < x < +\infty)$, 但 $\int_{-\infty}^{+\infty} p_1(x) dx = \int_0^{+\infty} \frac{1}{1+x^2} dx = \frac{\pi}{2} \neq 1$, 所以 $p_1(x)$ 不是概率密度函数.

(2) 当 $x \in (0, \pi)$ 时, $p_2(x) = \frac{1}{2} \cos x < 0$, 所以 $p_2(x)$ 不是概率密度函数.

(3) $p_3(x) \geq 0 (-\infty < x < +\infty)$, $\int_{-\infty}^{+\infty} p_3(x) dx = \int_a^{+\infty} e^{-(x-a)} dx = 1$, 所以 $p_3(x)$ 是概率密度函数.

例 2.2.2 设连续型随机变量 X 的概率密度函数为

$$p(x) = \begin{cases} c+x, & -1 < x \leq 0, \\ c-x, & 0 < x \leq 1, \\ 0, & \text{其他}, \end{cases} \quad \text{求: (1) 常数 } c; (2) P\left(-\frac{1}{2} < X \leq \frac{1}{2}\right).$$

解 (1) 因为 $\int_{-\infty}^{+\infty} p(x) dx = 1$, 所以 $\int_{-1}^0 (c+x) dx + \int_0^1 (c-x) dx = 1$, 解得 $c = 1$.

$$\begin{aligned} (2) P\left(-\frac{1}{2} < X \leq \frac{1}{2}\right) &= P\left(-\frac{1}{2} < X < \frac{1}{2}\right) = \int_{-\frac{1}{2}}^{\frac{1}{2}} p(x) dx \\ &= \int_{-\frac{1}{2}}^0 (1+x) dx + \int_0^{\frac{1}{2}} (1-x) dx = \frac{3}{4}. \end{aligned}$$

例 2.2.3 连续型随机变量 X 服从拉普拉斯分布, 其概率密度函数为 $p(x) = A e^{-|x|}$ ($-\infty < x < +\infty$) 求:

(1) 常数 A ;

(2) X 落在区间 $(0, 1)$ 内的概率;

(3) $P(X^2 < 1)$.

解 (1) 因为 $\int_{-\infty}^{+\infty} p(x) dx = 1$, 所以

$$\int_{-\infty}^{+\infty} p(x) dx = \int_{-\infty}^{+\infty} A e^{-|x|} dx = 2A \int_0^{+\infty} e^{-x} dx = 2A \Gamma(1) = 2A = 1, \text{ 解得 } A = \frac{1}{2}.$$

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

$$(3) P(X^2 < 1) = P(-1 < X < 1) = \int_{-1}^1 \frac{1}{2} e^{-|x|} dx = \frac{1}{2} \cdot 2 \int_0^1 e^{-x} dx = 1 - e^{-1}.$$

2. 利用连续型随机变量的概率密度计算概率

例 2.2.4 在数值计算中, 由于四舍五入引起的误差 X 服从均匀分布. 如果小数点后面第 5 位按四舍五入处理, 求误差在 0.000 03 和 0.000 06 之间的概率.

解 由题意知, 误差 X 在 $[-0.000\ 05, 0.000\ 05]$ 上服从均匀分布, 其概率分布密度函数为

$$p(x) = \begin{cases} 10\ 000, & -0.000\ 05 < x < 0.000\ 05, \\ 0, & \text{其他.} \end{cases}$$

$$\text{所以所求概率为 } P(0.000\ 03 \leq X \leq 0.000\ 06) = \int_{0.000\ 03}^{0.000\ 06} 10\ 000 dx = 0.3.$$

例 2.2.5 设随机变量 Y 在 $[0, 5]$ 上服从均匀分布, 求方程 $4x^2 + 4Yx + 2 = 0$ 有实根的概率.

解 在方程 $4x^2 + 4Yx + 2 = 0$ 中 x 是未知数, Y 是随机变量, 一旦它取定了值, 那么方程就是已知系数的二次方程. 问题是要使方程有实根而 Y 在它的取值范围内的概率有多大. 由题意知, Y 在 $[0, 5]$ 上服从均匀分布, 则其概率密度函数为

$$p(x) = \begin{cases} \frac{1}{5}, & 0 \leq x \leq 5, \\ 0, & \text{其他.} \end{cases}$$

方程有实根当且仅当 $\Delta \geq 0$, 即 $(4Y)^2 - 4 \times 4 \times 2 \geq 0$, $Y^2 \geq 2$, 所以所求概率为

$$\begin{aligned} P(Y^2 \geq 2) &= P(Y \geq \sqrt{2} \text{ 或 } Y \leq -\sqrt{2}) = P(Y \geq \sqrt{2}) + P(Y \leq -\sqrt{2}) \\ &= \int_{-\infty}^{-\sqrt{2}} P(x) dx + \int_{\sqrt{2}}^{+\infty} P(x) dx = \int_{\sqrt{2}}^5 \frac{1}{5} dx = 1 - \frac{\sqrt{2}}{5}. \end{aligned}$$

例 2.2.6 某仪器袋有 3 只独立工作的同型号电子元件, 其寿命都是服从参数 $\lambda = \frac{1}{600}$ 的指数分布. 求仪器在使用的最初 200 h 内至少有一只电子元件损坏的概率.

解 设 X 表示仪器在使用的最初 200 h 内 3 只元件中损坏的只数, 则 $X \sim B(3, p)$, 其中 p 表示在最初的 200 h 内每只元件损坏的概率. 另设 Y 表示每只元件的寿命, 则由题意知

$$Y \sim \Gamma\left(1, \frac{1}{600}\right) \text{ 其概率密度函数为 } p(y) = \begin{cases} \frac{1}{600} e^{-\frac{y}{600}}, & y > 0 \\ 0, & y \leq 0, \end{cases} \text{ 因此}$$

$$P = P(Y \leq 200) = \int_{-\infty}^{200} p(y) dy = \int_0^{200} \frac{1}{600} e^{-\frac{y}{600}} dy = 1 - e^{-\frac{1}{3}}.$$

所以 $X \sim B(3, 1 - e^{-\frac{1}{3}})$, 故所求概率为

$$P(X \geq 1) = 1 - P(X = 0) = 1 - C_3^0 (1 - e^{-\frac{1}{3}})^0 (e^{-\frac{1}{3}})^3 = 1 - e^{-1}.$$

注 本题牵涉到两个随机变量,其中一个分布的未知参数要借助另一个随机变量的概率密度函数求出,而解此题的关键是能够理解随机变量 X 服从二项分布.

例 2.2.7 设某类日光灯管的使用寿命服从参数 $\lambda = \frac{1}{2\,000}$ 的指数分布.

(1)任取一根灯管,求能正常使用 1 000 h 以上的概率;

(2)任取一根灯管,已正常使用了 1 000 h,求还能使用 1 000 h 以上的概率.

解 设 X 表示灯管的使用寿命,则 $X \sim \Gamma\left(1, \frac{1}{2\,000}\right)$,其概率密度函数为

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \lambda = \frac{1}{2\,000}.$$

$$(1) P(X > 1\,000) = 1 - P(X \leq 1\,000) = 1 - \int_0^{1\,000} \lambda e^{-\lambda x} dx$$

$$= 1 - (1 - e^{-1\,000\lambda}) = e^{-\frac{1}{2}} \approx 0.607.$$

$$\begin{aligned} (2) P(X > 2\,000 | X > 1\,000) &= \frac{P(X > 2\,000, X > 1\,000)}{P(X > 1\,000)} = \frac{P(X > 2\,000)}{P(X > 1\,000)} \\ &= \frac{1 - P(X \leq 2\,000)}{1 - P(X \leq 1\,000)} = e^{-\frac{1}{2}} \approx 0.607. \end{aligned}$$

注 由本题看出,任取一根灯管,能正常使用 1 000 h 以上的概率,与在已经正常使用了 1 000 h 的情况下,还能再正常使用 1 000 h 以上的概率相等,均为 0.607. 这是指数分布的一个有趣的“无记忆性”或“无后效性”,即只要 X 服从指数分布,则对任意的 $s > 0, t > 0$ 有

$$P(X > s + t | X > s) = P(X > t).$$

例 2.2.8 某公共汽车站从上午 7 时起每 15 min 发一班车,即在 7:00, 7:15, 7:30, ... 有汽车发出. 如果乘客到达此汽车站的时间 X 是在 7:00~7:30 内的均匀随机变量,求乘客在车站等候不到 5 min 的概率.

解 若将 7:00 作为时间的起点,则 X 服从区间 $[0, 30]$ 上的均匀分布,其概率密度函数为

$$p(x) = \begin{cases} \frac{1}{30}, & 0 \leq x \leq 30, \\ 0, & \text{其他.} \end{cases}$$

为了使等候的时间不到 5 min,乘客要在 7:10~7:15 之间或 7:25~

7:30 之间到汽车站,所以所求概率为

$$P(10 < X < 15) + P(25 < X < 30) = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}.$$

例 2.2.9 设 $p(x) = (ax^2 + bx + c)^{-1} (-\infty < x < +\infty)$,为使 $p(x)$ 为 $(-\infty, +\infty)$ 上的概率密度函数,系数 a, b, c 必须满足什么条件?

解 设 $h(x) = ax^2 + bx + c$.

(1)若 $p(x)$ 为概率密度函数,则 $p(x) \geq 0$,因此必有 $h(x) > 0$.

由 $h'(x) = 2ax + b, h''(x) = 2a$,当 $h''(x) = 2a > 0$ 时, $h(x)$ 有最小值 $c - \frac{b^2}{4a}$. 所以当且

仅当 $a > 0, c - \frac{b^2}{4a} > 0$ 时,

$$h(x) = ax^2 + bx + c > 0, p(x) = (ax^2 + bx + c)^{-1} > 0.$$

(2) 若 $p(x)$ 为概率密度函数, 则 $\int_{-\infty}^{+\infty} p(x) dx = 1$.

$$\begin{aligned} \text{而 } \int_{-\infty}^{+\infty} p(x) dx &= \int_{-\infty}^{+\infty} \frac{1}{a} \frac{d\left(x - \frac{b}{2a}\right)}{\left(x - \frac{b}{2a}\right)^2 + \left(\frac{\sqrt{4ac - b^2}}{2a}\right)^2} \\ &= \frac{2}{\sqrt{4ac - b^2}} \arctan \frac{2a}{\sqrt{4ac - b^2}} \left(x - \frac{b}{2a}\right) \Big|_{-\infty}^{+\infty} \\ &= \frac{2\pi}{\sqrt{4ac - b^2}}, \end{aligned}$$

因此, $4ac - b^2 = 4\pi^2$.

综上所述, 当 $a > 0, c - \frac{b^2}{4a} > 0, 4ac - b^2 = 4\pi^2$ 时, $p(x)$ 为 $(-\infty, +\infty)$ 上的概率密度函数.

同步练习 2.2

1. 连续型随机变量 X 服从马克思维尔分布, 其概率密度函数为

$$p(x) = \begin{cases} Ax^2 e^{-\frac{x^2}{b}}, & x > 0 \\ 0, & \text{其他}, \end{cases} (b > 0). \text{ 求常数 } A.$$

2. 设随机变量 X 的概率密度为 $p(x) = \frac{k}{e^{-x} + e^x} (-\infty < x < +\infty)$. 求:

(1) 常数 k ; (2) $P\left(0 < X < \frac{1}{2} \ln 3\right)$.

3. 设随机变量 Y 在区间 $[1, 6]$ 上服从均匀分布, 求方程 $x^2 + Yx + 1 = 0$ 有实根的概率.

4. 设随机变量 X 在 $[2, 6]$ 上服从均匀分布, 现对 X 进行三次独立观察, 求至少有两次观测值大于 3 的概率.

5. 设随机变量 X 的概率密度函数为 $p(x) = \begin{cases} 3x^2, & 0 < x < 1 \\ 0, & \text{其他}. \end{cases}$ Y 表示对 X 的三次独立重

复观察中事件 " $X \leq \frac{1}{2}$ " 出现的次数, 求:

(1) " $X \leq \frac{1}{2}$ " 至少出现一次的概率; (2) " $X \leq \frac{1}{2}$ " 恰好出现两次的概率.

6. 某市每小时消耗的电力 X 是服从参数为 $\alpha = 3, \beta = \frac{1}{2}$ 的 Γ 分布的随机变量, 假如每小时向该市供应 1 200 万度电力, 该城市仍电力不足的概率是多少?

7. 设带有 3 颗炸弹的轰炸机向敌方某铁路投弹, 某炸弹落在铁路两旁 40 m 以内, 将导致铁路交通受到破坏, 在一定投弹准确度下, 弹着点到铁路的距离 X 的密度函数为

$$p(x) = \begin{cases} \frac{100+x}{100\,000}, & -100 < x < 0, \\ \frac{100-x}{100\,000}; & 0 \leq x < 100, \\ 0, & \text{其他.} \end{cases}$$

若3颗炸弹全部使用,问敌方铁路交通受到破坏的概率是多大?

同步练习 2.2 参考答案

$$1. A = \frac{4}{b \sqrt{\pi b}}.$$

$$2. (1) k = \frac{2}{\pi}. \quad (2) P\left(0 < X < \frac{1}{2} \ln 3\right) = \frac{1}{6}.$$

$$3. P(Y \geq 2) + P(Y \leq -2) = \frac{4}{5}.$$

$$4. \text{设 } Y \text{ 表示三次独立观测中事件“} X > 3 \text{”出现的次数, } Y \sim B\left(3, \frac{3}{4}\right), P(Y \geq 2) = \frac{27}{32}.$$

$$5. (1) \frac{169}{512}, (2) \frac{21}{512}.$$

$$6. P(X > 12) = \int_{12}^{+\infty} \frac{1}{16} x^2 e^{-\frac{x}{2}} dx = 25e^{-6} \approx 0.0602.$$

$$7. P(-40 < X < 40) = 0.64,$$

$$\text{所求概率为 } 1 - (1 - 0.64)^3 = 0.9533.$$

2.3 随机变量的分布函数与随机变量函数的分布

2.3.1 知识要点

1. 分布函数

定义: 设 X 是一随机变量, 如果记 $F(x) = P(X \leq x)$, 则称 $F(x)$ 为随机变量 X 的分布函数.

性质: ① $0 \leq F(x) \leq 1 (-\infty < x < +\infty)$;

② $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$;

③ $F(x)$ 是 x 的不减函数, 即对于任意的二实数 $x_1 < x_2$, 总有 $F(x_1) \leq F(x_2)$.

2. 离散型随机变量的分布函数

设离散型随机变量 X 的概率分布为 $P(X = x_k) = p_k (k = 1, 2, \dots)$, 则其分布函数为 $F(x) = P(X \leq x) = \sum_{x_k \leq x} p_k$.

离散型随机变量 X 的分布函数的图形是一阶梯形曲线.

3. 连续型随机变量的分布函数

设连续型随机变量 X 的概率密度函数为 $p(x)$ 则其分布函数为 $F(x) = \int_{-\infty}^x p(t)dt$.

注 如果 $p(x)$ 在点 x 处连续 则有 $F'(x) = p(x)$.

4. 正态随机变量的概率计算

① 标准正态分布的分布函数 如果 $X \sim N(0, 1)$ 则其分布函数为

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (-\infty < x < +\infty).$$

② 标准正态分布的分布函数的性质:

1° $\Phi(-x) = 1 - \Phi(x)$;

2° 设 $X \sim N(0, 1)$ $x \geq 0$ 则有

$$P(|X| \geq x) = 2\Phi(-x), P(|X| \leq x) = 2\Phi(x) - 1;$$

3° 设 $X \sim N(\mu, \sigma^2)$ 则有

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

服从正态分布 $N(\mu, \sigma^2)$ 的随机变量 X 落在任意区间 (a, b) 中的概率 $P(a < X < b)$ 均可由上述 $\Phi(x)$ 的性质和标准正态分布表计算得出.

5. 随机变量函数的分布

① 随机变量的函数定义: 设 $f(x)$ 是定义在随机变量 X 一切可能取值集合上的函数, 如果当随机变量 X 取值 x 时, 随机变量 Y 就取值 $y = f(x)$, 则称 Y 为随机变量 X 的函数. 记为 $Y = f(X)$.

② 离散型随机变量函数的分布: 设 X 是离散型随机变量, 其概率分布为: $Y = f(x)$ 是随机变量 X 的函数.

X	x_1	x_2	x_3	$\dots$	x_k	$\dots$
$P(X = x_k)$	p_1	p_2	p_3	$\dots$	p_k	$\dots$

若 $f(x_1), f(x_2), f(x_3), \dots, f(x_k)$ 互不相等, 记 $y_i = f(x_i) (i = 1, 2, \dots)$, 则 Y 的概率分布为

Y	y_1	y_2	y_3	$\dots$	y_k	$\dots$
$P(Y = y_k)$	p_1	p_2	p_3	$\dots$	p_k	$\dots$

若 $f(x_1), f(x_2), f(x_3), \dots, f(x_k)$ 不是互不相等的, 应把那些相等的值分别合并, 并根据概率的加法公式把相应的 p_k 相加, 得到 Y 的概率分布.

③ 连续型随机变量函数的分布: 设连续型随机变量 X 的概率密度为 $p_X(x)$, 函数 $y = f(x)$ 处处可导, 且对任意的 x 有 $f'(x) > 0$ (或 $f'(x) < 0$) 则随机变量 X 的函数 $Y = f(X)$ 也是一个连续型随机变量, 其概率密度为.

$$p_Y(y) = \begin{cases} p_X(g(y)) \cdot |g'(y)|, & \alpha < y < \beta, \\ 0, & \text{其他.} \end{cases}$$

其中 $g(y)$ 是 $f(x)$ 的反函数,

$$\alpha = \min\{\lim_{x \rightarrow -\infty} f(x), \lim_{x \rightarrow +\infty} f(x)\}, \beta = \max\{\lim_{x \rightarrow -\infty} f(x), \lim_{x \rightarrow +\infty} f(x)\}.$$

2.3.2 典型题型解析

1. 有关分布函数性质的问题

例 2.3.1 下列函数中, 哪些可以作为随机变量的分布函数?

$$(1) F(x) = \frac{1}{1+x^2};$$

$$(2) F(x) = \frac{3}{4} + \frac{1}{2\pi} \arctan x;$$

$$(3) F(x) = e^{-e^{-x}}; \quad (4) F(x) = \begin{cases} 0, & x < -1, \\ \frac{1}{2}x^3 + \frac{1}{2}, & -1 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

解 (1) 因为 $\lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} \frac{1}{1+x^2} = 0 \neq 1$, 所以 $F(x)$ 不是分布函数.

(2) 因为 $\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} \left(\frac{3}{4} + \frac{1}{2\pi} \arctan x \right) = \frac{3}{4} + \frac{1}{2\pi} \left(-\frac{\pi}{2} \right) = \frac{1}{2} \neq 0$, 所以 $F(x)$ 不是分布函数.

(3) $F(x)$ 在 $(-\infty, +\infty)$ 内连续;

因为 $F'(x) = e^{-x} \cdot e^{-e^{-x}} > 0$, 所以 $F(x)$ 在 $(-\infty, +\infty)$ 内单调增加;

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

综上所述, $F(x) = e^{-e^{-x}}$ 可以作为随机变量的分布函数.

$$(4) \lim_{x \rightarrow +\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1;$$

$$F'(x) = \begin{cases} 0, & x < -1, \\ \frac{3}{2}x^2, & -1 < x < 1, \\ 0, & x \geq 1. \end{cases}$$

在 $(-1, 1)$ 内 $F'(x) > 0$, 所以 $F(x)$ 在 $(-1, 1)$ 内单调增加; 且

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

综上所述, $F(x)$ 可以作为随机变量的分布函数.

注 解此类题的关键是, 检验函数 $F(x)$ 是否满足分布函数的基本性质.

2. 已知概率分布(或概率密度)求分布函数

例 2.3.2 已知随机变量 X 的概率分布为

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

求 (1) X 的分布函数;

$$(2) P\left(X < \frac{1}{2}\right), P\left(1 < X \leq \frac{2}{3}\right), P(2 \leq x \leq 3).$$

$$\text{解 } (1) F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{6}, & -1 \leq x < 2 \\ \frac{1}{6} + \frac{2}{6}, & 2 \leq x < 3 \\ \frac{1}{6} + \frac{2}{6} + \frac{3}{6}, & x \geq 3 \end{cases} = \begin{cases} 0, & x < -1, \\ \frac{1}{6}, & -1 \leq x < 2, \\ \frac{1}{2}, & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

$$(2) P\left(X < \frac{1}{2}\right) = F\left(\frac{1}{2} - 0\right) = \sum_{x_k < \frac{1}{2}} P_k = \frac{1}{6},$$

$$P\left(1 < x \leq \frac{2}{3}\right) = P\left(X \leq \frac{3}{2}\right) - P(X \leq 1) = F\left(\frac{3}{2}\right) - F(1) = \frac{1}{6} - \frac{1}{6} = 0.$$

$$P(2 \leq X \leq 3) = P(X \leq 3) - P(X < 2) = F(3) - F(2 - 0) = 1 - \frac{1}{6} = \frac{5}{6}.$$

注 (1) 设离散型随机变量 X 的分布为 $p_k = P(X = x_k)$, $k = 1, 2, \dots$. 则 X 的分布函数为

$$F(x) = P(X \leq x) = P(X \text{ 的取值不大于 } x \text{ 的一切 } x_k)$$

$$= \sum_{x_k \leq x} P(X = x_k) = \sum_{x_k \leq x} p_k.$$

将此式写成分段函数的形式为

$$F(x) = \begin{cases} 0, & x < x_1, \\ p_1, & x_1 \leq x < x_2, \\ p_1 + p_2, & x_2 \leq x < x_3, \\ \vdots & \vdots \\ p_1 + p_2 + \dots + p_{n-1}, & x_{n-1} \leq x < x_n, \\ p_1 + p_2 + \dots + p_{n-1} + p_n = 1, & x \geq x_n. \end{cases}$$

(2) n 个常用等式

$$P(X > x) = 1 - F(x),$$

$$P(X = x) = F(x) - F(x - 0),$$

$$P(X < x) = F(x - 0),$$

$$P(X \geq x) = 1 - F(x - 0),$$

$$P(a < X \leq b) = F(b) - F(a),$$

$$P(a < X < b) = F(b - 0) - F(a),$$

$$P(a \leq X \leq b) = F(b) - F(a - 0),$$

$$P(a \leq X < b) = F(b - 0) - F(a - 0).$$

例 2.3.3 设 X 的概率密度为 $f(x) = \begin{cases} \frac{1}{2} \cos x, & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}, \\ 0, & \text{其他.} \end{cases}$

求 (1) 分布函数 $F(x)$;

$$(2) P\left(\frac{\pi}{4} < X < \frac{\pi}{3}\right)$$

解 (1) 当 $x < -\frac{\pi}{2}$ 时, $F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x 0 dt = 0$;

当 $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$ 时, $F(x) = \int_{-\infty}^x f(t) dt = \int_{-\frac{\pi}{2}}^x \frac{1}{2} \cos t dt = \frac{1}{2} (1 + \sin x)$;

当 $x > \frac{\pi}{2}$ 时, $F(x) = \int_{-\infty}^x f(t) dt = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2} \cos t dt = 1$.

所以分布函数为

$$F(x) = \begin{cases} 0, & x < -\frac{\pi}{2}, \\ \frac{1}{2}(1 + \sin x), & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}, \\ 1, & x > \frac{\pi}{2}. \end{cases}$$

$$\begin{aligned} (2) P\left(\frac{\pi}{4} < X < \frac{\pi}{3}\right) &= F\left(\frac{\pi}{3}\right) - F\left(\frac{\pi}{4}\right) = \frac{1}{2}\left(1 + \sin \frac{\pi}{3}\right) - \frac{1}{2}\left(1 + \sin \frac{\pi}{4}\right) \\ &= \frac{1}{4}(\sqrt{3} - \sqrt{2}). \end{aligned}$$

注 (1) 设连续型随机变量 X 的概率密度函数为 $p(x)$, 则其分布函数为

$$F(x) = \int_{-\infty}^x p(t) dt.$$

(2) 如果 X 为连续型随机变量, 概率密度函数为 $p(x)$, 则

$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b p(x) dx \\ &= F(b) - F(a); \end{aligned}$$

$$P(X \leq b) = P(X < b) = \int_{-\infty}^b p(x) dx;$$

$$P(X \geq a) = P(X > a) = 1 - P(X \leq a) = 1 - \int_{-\infty}^a p(x) dx.$$

3. 已知分布函数求概率分布(或概率密度)

例 2.3.4 设 X 的分布函数为

$$F(x) = \begin{cases} 0, & x < -1, \\ 0.3, & -1 \leq x < 0, \\ 0.6, & 0 \leq x < 1, \\ 0.8, & 1 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

求 (1) X 的概率分布;

$$(2) P\left(\frac{5}{2} < X < \frac{7}{2}\right).$$

解 (1) X 的可能取值是 $-1, 0, 1, 3$, 且 $p(X = x_k) = F(x_k) - F(x_k - 0)$,

$$P(X = -1) = F(-1) - F(-1 - 0) = 0.3 - 0 = 0.3,$$

$$P(X = 0) = F(0) - F(0 - 0) = 0.6 - 0.3 = 0.3,$$

$$P(X = 1) = F(1) - F(1 - 0) = 0.8 - 0.6 = 0.2,$$

$$P(X=3)=F(3)-F(3-0)=1-0.8=0.2.$$

所以 X 的分布为

X	-1	0	1	3
P	0.3	0.3	0.2	0.2

$$(2) \text{解法 1 } P\left(\frac{5}{2} < X < \frac{7}{2}\right) = P(X=3) = 0.2$$

$$\text{解法 2 } P\left(\frac{5}{2} < X < \frac{7}{2}\right) = F\left(\frac{7}{2}-0\right) - F\left(\frac{5}{2}\right) = 1 - 0.8 = 0.2.$$

注 已知离散型随机变量 X 的分布函数 $F(x)$ 求 X 的概率分布的方法有:

① X 的可能取值是 $F(x)$ 的所有分段点 $x_k (k=1, 2, \dots)$;

② $P(X=x_k) = F(x_k) - F(x_k-0) (k=1, 2, \dots)$.

例 2.3.5 设连续型随机变量 X 的分布函数为

$$F(x) = \begin{cases} 0, & x < -a, \\ A + B \arcsin \frac{x}{a}, & -a \leq x < a, \text{ 这里 } a > 0. \\ 1, & x \geq a, \end{cases}$$

求 (1) 常数 A 和 B ;

(2) 概率密度函数 $p(x)$.

解 (1) 因为 X 是连续型随机变量, 所以 $F(x)$ 连续, 在 $x = -a$ 与 $x = a$ 处也连续. 因此有

$$\lim_{x \rightarrow a^+} F(x) = \lim_{x \rightarrow -a^-} F(x) = F(-a), \lim_{x \rightarrow a^-} F(x) = F(a),$$

即 $A - \frac{\pi}{2}B = 0$, $A + \frac{\pi}{2}B = 1$, 解得 $A = \frac{1}{2}$, $B = \frac{1}{\pi}$.

$$(2) p(x) = F'(x) = \begin{cases} 0, & x < -a, \\ \frac{1}{\pi \sqrt{a^2 - x^2}}, & -a < x < a, \\ 0, & x > a. \end{cases}$$

当 $x = a$ 时

$$\begin{aligned} \lim_{x \rightarrow a^+} \frac{F(x) - F(a)}{x - a} &= \lim_{x \rightarrow a^+} \frac{1 - 1}{x - a} = 0; \\ \lim_{x \rightarrow a^-} \frac{F(x) - F(a)}{x - a} &= \lim_{x \rightarrow a^-} \frac{\frac{1}{2} + \frac{1}{\pi} \arcsin \frac{x}{a} - 1}{x - a} \\ &= \lim_{x \rightarrow a^-} \frac{\frac{1}{\pi} \arcsin \frac{x}{a} - \frac{1}{2}}{x - a} \\ &\stackrel{\left(\frac{0}{0}\right)}{=} \lim_{x \rightarrow a^-} \frac{1}{\pi \sqrt{a^2 - x^2}} = \infty. \end{aligned}$$

因此 $F(x)$ 在点 $x = a$ 处不可导, 同理, $F(x)$ 在 $x = -a$ 处也不可导. 在这两点上可补充定义: $p(a) = p(-a) = 0$, 在这两点处的函数值并不影响 $p(x)$ 在区间上的积分值, 所以 X 的概率密度为

$$p(x) = \begin{cases} \frac{1}{\pi\sqrt{a^2 - x^2}}, & -a < x < a, \\ 0, & \text{其他.} \end{cases}$$

注 已知连续型随机变量 X 的分布函数 $F(x)$ 求其概率密度函数. 可直接由 $F(x)$ 按求导公式和法则求导, 即 $p(x) = F'(x)$, 在分段点处一律令 $p(x) = 0$.

4. 有关正态分布的计算题

例 2.3.6 设 $X \sim N(\mu, \sigma^2)$, 求:

(1) $P(a < X < b)$;

(2) $P(\mu - \sigma < X < \mu + \sigma)$;

(3) $P(\mu - 2\sigma < X < \mu + 2\sigma)$;

(4) $P(\mu - 3\sigma < X < \mu + 3\sigma)$.

解 (1) $P(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$.

设 $\frac{x-\mu}{\sigma} = t$, $x = \mu + \sigma t$, $dx = \sigma dt$, 则有

$$P(a < X < b) = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

查标准正态分布表, 就可求出这个概率值.

利用(1)题结论, 可得:

(2) $P(\mu - \sigma < X < \mu + \sigma) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826$.

(3) $P(\mu - 2\sigma < X < \mu + 2\sigma) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0.9544$.

(4) $P(\mu - 3\sigma < X < \mu + 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 0.9974$.

例 2.3.7 设 $X \sim N(10, \sigma^2)$, $P(10 < X < 20) = 0.3$, 求 $P(0 < X < 10)$

解法 1 因为 $P(10 < X < 20) = \Phi\left(\frac{20-10}{\sigma}\right) - \Phi\left(\frac{10-10}{\sigma}\right) = \Phi\left(\frac{10}{\sigma}\right) - \Phi(0)$
 $= \Phi\left(\frac{10}{\sigma}\right) - \frac{1}{2} = 0.3$, 所以 $\Phi\left(\frac{10}{\sigma}\right) = 0.8$.

而 $P(0 < X < 10) = \Phi\left(\frac{10-10}{\sigma}\right) - \Phi\left(\frac{0-10}{\sigma}\right) = \Phi(0) - \Phi\left(-\frac{10}{\sigma}\right)$
 $= \Phi(0) + \Phi\left(\frac{10}{\sigma}\right) - 1 = \frac{1}{2} + 0.8 - 1 = 0.3$.

解法 2 因为 $X \sim N(10, \sigma^2)$, 所以 $p(x)$ 曲线关于直线 $x = \mu = 10$ 对称, 所以在 $[0, 10]$ 与 $[10, 20]$ 上 X 取值的概率相等, 因此 $P(0 < X < 10) = P(10 < X < 20) = 0.3$.

例 2.3.8 设 $X \sim N(108, 9)$, 求:

(1) $P(101.1 < X < 117.6)$;

(2) 求常数 a , 使 $P(X < a) = 0.90$;

(3) 求常数 a , 使 $P(|X - a| > a) = 0.01$.

解 (1) $P(101.1 < X < 117.6) = \Phi\left(\frac{117.6-108}{3}\right) - \Phi\left(\frac{101.1-108}{3}\right)$
 $= \Phi(3.2) - \Phi(-2.3) = \Phi(3.2) + \Phi(2.3) - 1 = 0.9995 + 0.9893 - 1 = 0.9888$.

$$(2) P(X < a) = \Phi\left(\frac{a-108}{3}\right) = 0.90.$$

查表知 $\Phi(1.29) \approx 0.90$ 所以 $\frac{a-108}{3} = 1.29$ 即 $a = 112.17$.

$$(3) P(|X-a| > a) = P(X > 2a) + P(X < 0) = 1 - P(X \leq 2a) + P(X < 0) \\ = 1 - \Phi\left(\frac{2a-108}{3}\right) = 0.01,$$

即 $\Phi\left(\frac{2a-108}{3}\right) = 0.99.$

查表得 $\frac{2a-108}{3} \approx 2.33$ 所以 $a = 57.4$.

例 2.3.9 设随机变量 $X \sim N(0, 100)$ 求在 100 次独立试验中, 至少有 3 次 X 的绝对值大于 19.6 的概率.

解 设随机变量 Y 表示 100 次独立试验中 X 的绝对值大于 19.6 的次数, 则 $Y \sim B(100, p)$ 其中 p 为每次试验中 X 的绝对值大于 19.6 的概率.

因为 $X \sim N(0, 100)$ 所以

$$P(|X| > 19.6) = P(X > 19.6) + p(X < -19.6) \\ = 1 - F(19.6) + F(-19.6) = 1 - \Phi(1.96) + \Phi(-1.96) \\ = 2 - 2\Phi(1.96) = 2 - 2 \times 0.9750 = 0.05.$$

即 $Y \sim B(100, 0.05)$ 因此

$$p(Y \geq 3) = 1 - p(Y < 3) \\ = 1 - P(Y=0) - P(Y=1) - P(Y=2) \\ = 1 - 0.95^{100} - C_{100}^1 (0.95)^{99} (0.05)^1 - C_{100}^2 (0.95)^{98} (0.05)^2.$$

由于 n 很大, p 很小, $np = \lambda = 5$ 所以由 Poisson 定理有

$$P(Y \geq 3) = 1 - P(Y < 3) = 1 - P(Y=0) - P(Y=1) - P(Y=2) \\ \approx 1 - e^{-5} - 5e^{-5} - \frac{25}{2}e^{-5} = 0.87.$$

注 在正态随机变量的概率计算中, 经常用到下列几个重要等式:

$$\textcircled{1} P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right);$$

$$\textcircled{2} P(X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right);$$

$$\textcircled{3} P(X > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right);$$

$$\textcircled{4} \Phi(-x) = 1 - \Phi(x).$$

5. 有关随机变量函数分布的计算题

例 2.3.10 设随机变量 X 的概率分布为

X	-2	0	2	3
P	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

求 (1) $Y = 3X$ 的概率分布;

(2) $Y = X^2$ 的概率分布.

解 (1) $Y = 3X$ 的概率分布为

Y	-6	0	6	9
P	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

(2) $Y = X^2$ 的概率分布为

Y	0	4	9
P	$\frac{1}{7}$	$\frac{4}{7}$	$\frac{2}{7}$

例 2.3.11 设随机变量 X 的概率分布为

X	1	2	...	n	...
P	$\frac{1}{2}$	$\frac{1}{2^2}$	...	$\frac{1}{2^n}$	...

求 $Y = \sin\left(\frac{\pi}{2}X\right)$ 的概率分布.

解 因为 $\sin\left(\frac{n}{2}\pi\right) = \begin{cases} -1, & n=4k-1, \\ 0, & n=2k, \text{ 其中 } k=1, 2, \dots \\ 1, & n=4k-3, \end{cases}$

所以 $Y = \sin\left(\frac{\pi}{2}X\right)$ 有三个可能值 $-1, 0, 1$

$$P(Y = -1) = P(X = 3) + P(X = 7) + P(X = 11) + \dots$$

$$= \frac{1}{2^3} + \frac{1}{2^7} + \frac{1}{2^{11}} + \dots = \frac{\frac{1}{2^3}}{1 - \frac{1}{2^4}} = \frac{2}{15}.$$

$$P(Y = 0) = P(X = 2) + P(X = 4) + P(X = 6) + \dots$$

$$= \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \dots = \frac{\frac{1}{2^2}}{1 - \frac{1}{2^2}} = \frac{1}{3}.$$

$$P(Y = 1) = P(X = 1) + P(X = 5) + P(X = 9) + \dots$$

$$= \frac{1}{2} + \frac{1}{2^5} + \frac{1}{2^9} + \dots = \frac{\frac{1}{2}}{1 - \frac{1}{2^4}} = \frac{8}{15}.$$

所以 $Y = \sin\left(\frac{\pi}{2}X\right)$ 的概率分布为

Y	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
P	$\frac{2}{15}$	$\frac{1}{3}$	$\frac{8}{15}$

注 在求离散型随机变量函数的分布时, 关键是通过 $Y = f(X)$ 找出 Y 所有可能的取值, 然后分别通过对应的 X 的取值来确定 Y 取这些值的概率.

例 2.3.12 已知随机变量 $X \sim N(0, 1)$. 求 $Y = -X$ 的概率密度函数.

解法 1 函数 $y = -x$ 在 $(-\infty, +\infty)$ 上单调可导, 反函数为 $x = -y, \frac{dx}{dy} = -1 \neq 0$. 所以 Y 的概率密度函数为

$$\begin{aligned} p_Y(y) &= \varphi(-y) \cdot |-1| = \varphi(-y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-y)^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad (-\infty < y < +\infty). \end{aligned}$$

解法 2 $F_Y(y) = P(Y \leq y) = P(-X \leq y) = P(X \geq -y)$

$$= 1 - \Phi(-y) = \Phi(y).$$

$$\text{所以 } p_y(y) = F'_Y(y) = \Phi'(y) = \varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad (-\infty < y < +\infty).$$

例 2.3.13 设随机变量 X 在区间 $[0, \pi]$ 上服从均匀分布, 求 $Y = \sin X$ 的概率密度函数.

解 因为 $y = \sin x$ 在 $[0, \pi]$ 上不单调, 所以只能用分布函数法先求 $F_Y(y)$, 再求导得 $p_Y(y)$.

当 $y < 0$ 时, $F_Y(y) = P(Y \leq y) = 0$.

$$\begin{aligned} \text{当 } 0 \leq y < 1 \text{ 时, } F_Y(y) &= P(Y \leq y) = P(\sin X \leq y) = \int_0^{\arcsin y} \frac{1}{\pi} dx + \int_{\pi - \arcsin y}^{\pi} \frac{1}{\pi} dx \\ &= \frac{2}{\pi} \arcsin y. \end{aligned}$$

当 $y \geq 1$ 时, $F_Y(y) = P(Y \leq y) = P(\sin X \leq y) = 1$.

对分布函数求导得到 Y 的概率密度函数

$$p_Y(y) = \begin{cases} \frac{2}{\pi \sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{其他.} \end{cases}$$

例 2.3.14 设随机变量 X 的概率密度为 $p_X(x)$, 求下列随机变量 Y 的概率密度:

$$(1) Y = X^2 \quad (2) Y = |X| \quad (3) Y = \frac{1}{X}$$

解 (1) 因为 $y = x^2$ 在 $(-\infty, +\infty)$ 内不是单调函数, 所以用分布函数法.

当 $y \leq 0$ 时, $F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(\emptyset) = 0$;

$$\begin{aligned} \text{当 } y > 0 \text{ 时, } F_Y(y) &= P(Y \leq y) = P(x^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

所以 $Y = X^2$ 的概率密度函数为

$$p_Y(y) = \begin{cases} \frac{p_X(\sqrt{y}) + p_X(-\sqrt{y})}{2\sqrt{y}}, & y > 0, \\ 0, & \text{其他.} \end{cases}$$

(2) 当 $y \leq 0$ 时, $F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(\emptyset) = 0$.

当 $y > 0$ 时, $F_Y(y) = P(Y \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y)$.

所以 $Y = |X|$ 的概率密度函数为

$$p_Y(y) = \begin{cases} p_X(y) + p_X(-y), & y > 0, \\ 0, & \text{其他.} \end{cases}$$

(3) 当 $y > 0$ 时, $F_Y(y) = P(Y \leq y) = P\left(\frac{1}{X} \leq y\right) + P\left(0 < \frac{1}{X} \leq y\right)$

$$= P(X < 0) + P\left(X \geq \frac{1}{y}\right)$$

$$= F_X(0) + 1 - F_X\left(\frac{1}{y}\right).$$

当 $y < 0$ 时, $F_Y(y) = P(Y \leq y) = P\left(\frac{1}{y} \leq X < 0\right) = F_X(0) - F_X\left(\frac{1}{y}\right)$.

当 $y = 0$ 时, $F_Y(y) = P(Y \leq y) = P\left(\frac{1}{X} \leq 0\right) = P(X < 0) = F_X(0)$.

所以 $Y = \frac{1}{x}$ 的概率密度函数为

$$p_Y(y) = \begin{cases} \frac{1}{y^2} p_X\left(\frac{1}{y}\right), & y \neq 0, \\ 0, & y = 0. \end{cases}$$

例 2.3.15 设随机变量 X 在 $[1, 2]$ 上服从均匀分布. 求:

(1) $Y = e^{2X}$ 的概率密度函数;

(2) $Y = -2\ln X$ 的概率密度函数.

解 (1) $y = e^{2x}$, $y' = 2e^{2x} > 0$, $x = g(y) = \frac{1}{2} \ln y$, $g'(y) = \frac{1}{2y}$,

$$\alpha = \min \left\{ \lim_{x \rightarrow 1^+} e^{2x}, \lim_{x \rightarrow 2^-} e^{2x} \right\} = \min \{e^2, e^4\} = e^2,$$

$$\beta = \max \left\{ \lim_{x \rightarrow 1^+} e^{2x}, \lim_{x \rightarrow 2^-} e^{2x} \right\} = \max \{e^2, e^4\} = e^4,$$

所以由公式可得 $Y = e^{2X}$ 的概率密度函数为

$$p_Y(y) = \begin{cases} p_X(g(y)) \cdot |g'(y)|, & \alpha \leq y \leq \beta, \\ 0, & \text{其他} \end{cases}$$

$$= \begin{cases} \frac{1}{2y}, & e^2 \leq y \leq e^4, \\ 0, & \text{其他}. \end{cases}$$

(2) $y = -2\ln x$, $y' = -\frac{2}{x} < 0$ ($1 \leq x \leq 2$), $x = g(y) = e^{-\frac{y}{2}}$, $g'(y) = -\frac{1}{2} e^{-\frac{y}{2}}$,

$$\alpha = \min \left\{ \lim_{x \rightarrow 1^+} (-2\ln x), \lim_{x \rightarrow 2^-} (-2\ln x) \right\} = \min \left\{ 0, \ln \frac{1}{4} \right\} = \ln \frac{1}{4},$$

$$\beta = \max \left\{ \lim_{x \rightarrow 1^+} (-2\ln x), \lim_{x \rightarrow 2^-} (-2\ln x) \right\} = \max \left\{ 0, \ln \frac{1}{4} \right\} = 0,$$

所以由公式可得 $Y = -2\ln X$ 的概率密度函数为

$$p_Y(y) = \begin{cases} p_X(g(y)) \cdot |g'(y)|, & \alpha \leq y \leq \beta, \\ 0, & \text{其他} \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{-\frac{y}{2}}, & \ln \frac{1}{4} \leq y \leq 0, \\ 0, & \text{其他}. \end{cases}$$

注 求连续型随机变量函数的概率密度函数有如下两种方法.

方法一: 公式法.

如果函数 $y = f(x)$ 是单调可导函数, 且 $f'(x) \neq 0$, 则 $Y = f(X)$ 的概率密度函数为

$$p_Y(y) = \begin{cases} p_X(g(y)) \cdot |g'(y)|, & \alpha < y < \beta, \\ 0, & \text{其他}. \end{cases}$$

其中 $g(y)$ 是 $f(x)$ 的反函数, $\alpha = \min \left\{ \lim_{x \rightarrow -\infty} f(x), \lim_{x \rightarrow +\infty} f(x) \right\}$,

$$\beta = \max \{ \lim_{x \rightarrow -\infty} f(x), \lim_{x \rightarrow +\infty} f(x) \}.$$

特别地, 如果 $p_X(x)$ 在有限区间 $[a, b]$ 以外等于零, 则

$$\alpha = \min \{ \lim_{x \rightarrow a^+} f(x), \lim_{x \rightarrow b^-} f(x) \},$$

$$\beta = \max \{ \lim_{x \rightarrow a^+} f(x), \lim_{x \rightarrow b^-} f(x) \}.$$

方法二: 分布函数法.

先求 Y 的分布函数

$$F_Y(y) = P(Y \leq y) = P(f(x) \leq y).$$

再对上式两边求导, 即可求出 Y 的概率密度函数 $p_Y(y)$.

例 2.3.16 对圆的直径做近似测量, 设其值均匀分布在区间 $[a, b]$ 内, 求圆面积的概率密度函数.

解 设随机变量 X 表示圆的直径, Y 表示圆的面积, 则 $Y = \frac{1}{4}\pi X^2$.

$$X \text{ 的概率密度函数为 } p_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{其他}, \end{cases} (a > 0, b > 0).$$

$$y = \frac{1}{4}\pi x^2, x' = \frac{1}{2}\pi x > 0 (a \leq x \leq b), x = g(y) = \sqrt{\frac{4y}{\pi}},$$

$$x' = g'(y) = \frac{1}{\sqrt{\pi y}}, \alpha = \min \left\{ \lim_{x \rightarrow a^+} \frac{1}{4}\pi x^2, \lim_{x \rightarrow b^-} \frac{1}{4}\pi x^2 \right\} = \frac{1}{4}\pi a^2,$$

$$\beta = \max \left\{ \lim_{x \rightarrow a^+} \frac{1}{4}\pi x^2, \lim_{x \rightarrow b^-} \frac{1}{4}\pi x^2 \right\} = \frac{1}{4}\pi b^2,$$

所以由公式可得圆面积 Y 的概率密度函数为

$$p_Y(y) = \begin{cases} p_X(g(y)) \cdot |g'(y)|, & \alpha \leq y \leq \beta, \\ 0, & \text{其他} \end{cases}$$

$$= \begin{cases} \frac{1}{(b-a)\sqrt{\pi y}}, & \frac{1}{4}\pi a^2 \leq y \leq \frac{1}{4}\pi b^2, \\ 0, & \text{其他}. \end{cases}$$

例 2.3.17 已知随机变量 X 的概率密度为 $p(x) = \frac{2}{\pi} \frac{1}{e^x + e^{-x}} (-\infty < x < +\infty)$.

求 $Y = f(X)$ 的概率分布, 其中 $f(x) = \begin{cases} -1, & x < 0, \\ 1, & x \geq 0. \end{cases}$

解 Y 的可能取值为 $-1, 1$,

$$P(Y = -1) = P(X < 0) = \int_{-\infty}^0 \frac{2}{\pi} \frac{1}{e^x + e^{-x}} dx = \frac{2}{\pi} \int_{-\infty}^0 \frac{e^x}{1 + e^{2x}} dx$$

$$= \frac{2}{\pi} \arctan e^x \Big|_{-\infty}^0 = \frac{1}{2},$$

$$P(Y = 1) = P(X \geq 0) = \int_0^{+\infty} \frac{2}{\pi} \frac{1}{e^x + e^{-x}} dx = \frac{2}{\pi} \int_0^{+\infty} \frac{e^x}{1 + e^{2x}} dx$$

$$= \frac{2}{\pi} \arctan e^x \Big|_0^{+\infty} = \frac{1}{2}.$$

所以 Y 的概率分布为

Y	-1	1
P	$\frac{1}{2}$	$\frac{1}{2}$

注 此题说明,连续型随机变量的函数可以是离散型随机变量.

同步练习 2.3

1. 下列四个函数,哪个不能作为随机变量的分布函数?

$$\begin{aligned}
 (1) F_1(x) &= \begin{cases} 0, & x < 0, \\ \frac{x^2}{2}, & 0 \leq x < 1, \\ 1, & x \geq 1; \end{cases} & (2) F_2(x) &= \begin{cases} 0, & x < 0, \\ \frac{1}{3}, & 0 \leq x < 1, \\ \frac{1}{2}, & 1 \leq x < 2, \\ 1, & x \geq 2; \end{cases} \\
 (3) F_3(x) &= \begin{cases} \frac{\ln(1+x)}{1+x}, & x \geq 1, \\ 0, & x < 1; \end{cases} & (4) F_4(x) &= \begin{cases} 1 - e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}
 \end{aligned}$$

2. 袋中有标号为 $-1, 1, 1, 2, 2, 2$ 的 6 个球,从中任取一球,设 X 表示所得球的标号.求:

(1) X 的概率分布;

(2) X 的分布函数 $F(x)$;

(3) $P\left(X \leq \frac{1}{2}\right), P\left(1 < X \leq \frac{3}{2}\right), P\left(1 \leq X \leq \frac{3}{2}\right).$

3. 设随机变量 X 的分布函数为 $F(x) = \begin{cases} 0, & x < -1, \\ \frac{1}{4}, & -1 \leq x < 0, \\ \frac{3}{4}, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$

求 X 的概率分布.

4. 设连续型随机变量 X 的分布函数为

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

求 (1) 常数 A, B, C 的值;

(2) X 的概率密度函数;

(3) $P\left(X > \frac{1}{2}\right).$

5. 设连续型随机变量 X 的概率密度为 $p(x) = \begin{cases} Ax^2 e^{-2x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$

求 (1) 常数 A 的值;

(2) X 的分布函数;

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

6. 由某机器生产的螺栓的长度服从参数 $\mu = 10.05$, $\sigma = 0.06$ 的正态分布, 规定长度在 10.05 ± 0.12 内为合格. 求一螺栓为不合格品的概率.

7. 某高校抽样调查结果表明, 考生的外语成绩服从参数 $\mu = 72$ 的正态分布, 96 分以上的考生占总数的 2.3%. 求考生的外语成绩在 60 分至 84 分之间的概率.

8. 设连续型随机变量 $X \sim N(3, 2^2)$.

(1) 求 $P(|X| > 2)$;

(2) 决定 c , 使得 $P(X > c) = P(X \leq c)$.

9. 设离散型随机变量的概率分布为

X	-2	$-\frac{1}{2}$	0	2	4
P	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{3}$

求 (1) $X+2$ (2) $-X+1$ (3) X^2 的概率分布.

10. 设随机变量 $X \sim N(0, 1)$. 求 $Y = |X|$ 的概率密度.

11. 随机变量 X 的概率密度函数为 $p_X(x) = \begin{cases} 3x^2, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$, 求 $Y = \frac{1}{X}$ 的分布函数

$F_Y(y)$ 和概率密度.

12. 设随机变量 X 的概率密度 $p_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{其他} \end{cases}$, 求 $Y = 3X + 1$ 的概率密度.

13. 设随机变量 X 的概率密度为 $p_X(x) = \begin{cases} \frac{2}{\pi^2}x, & 0 < x < \pi \\ 0, & \text{其他} \end{cases}$, 求 $Y = \sin X$ 的概率密度.

14. 设电流 I 是一个随机变量, 它均匀分布在 $9 \sim 11$ A 之间, 此电流通过 2Ω 的电阻时消耗的功率 $W = 2I^2$. 求 W 的概率密度.

同步练习 2.3 参考答案

1. $F_3(x)$.

2. (1)

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

.

(2) $F(x) = \begin{cases} 0, & x < -1, \\ \frac{1}{6}, & -1 \leq x < 1, \\ \frac{1}{2}, & 1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$

(3) $P\left(X \leq \frac{1}{2}\right) = \frac{1}{6}$, $P\left(1 < X \leq \frac{3}{2}\right) = 0$, $P\left(1 \leq X \leq \frac{3}{2}\right) = \frac{1}{3}$.

$$3. \begin{array}{c|ccc} X & -1 & 0 & 1 \\ \hline P & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}.$$

$$4. (1) A=0, B=\frac{1}{2}, C=2.$$

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

$$(3) P\left(X > \frac{1}{2}\right) = \frac{7}{8}.$$

$$5. (1) \text{因为} \int_{-\infty}^{+\infty} p(x) dx = \int_0^{+\infty} A x^2 e^{-2x} dx = \frac{2A\Gamma(2)}{8} = 1, \text{所以 } A=4.$$

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

$$(3) P\left(0 < X < \frac{1}{2}\right) = F\left(\frac{1}{2}\right) - F(0) = 1 - \frac{5}{2e}.$$

$$\begin{aligned} 6. P &= 1 - P(10.05 - 0.12 < X < 10.05 + 0.12) \\ &= 1 - \Phi(2) + \Phi(-2) \\ &= 0.0456. \end{aligned}$$

$$7. \sigma = 12, P(60 \leq X \leq 84) = \Phi(1) - \Phi(-1) = 0.682.$$

$$8. (1) P(|X| > 2) = 1 - P(|X| \leq 2) = 1 - \left(\Phi\left(-\frac{1}{2}\right) - \Phi\left(-\frac{5}{2}\right)\right) = 0.6927.$$

$$(2) c = 3.$$

$$9. (1) \begin{array}{c|ccccc} X+2 & 0 & \frac{3}{2} & 2 & 4 & 6 \\ \hline P & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{6} & \frac{1}{3} \end{array}.$$

$$(2) \begin{array}{c|ccccc} -X+1 & 3 & \frac{3}{2} & 1 & -1 & -3 \\ \hline P & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{6} & \frac{1}{3} \end{array}.$$

$$(3) \begin{array}{c|cccc} X^2 & 0 & \frac{1}{4} & 4 & 16 \\ \hline P & \frac{1}{8} & \frac{1}{4} & \frac{7}{24} & \frac{1}{3} \end{array}.$$

$$10. p_Y(y) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}y^2}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

$$11. (1) F_Y(y) = \begin{cases} 0, & y < 0, \\ 0, & 0 < y < 1, \\ 1 - \frac{1}{y^3}, & y > 1 \end{cases} = \begin{cases} 1 - \frac{1}{y^3}, & y > 1, \\ 0, & \text{其他.} \end{cases}$$

$$(2) p_Y(y) = \begin{cases} \frac{3}{y^4}, & y > 1, \\ 0, & \text{其他.} \end{cases}$$

$$12. p_Y(y) = \begin{cases} \frac{1}{3}, & 1 \leq y \leq 4, \\ 0, & \text{其他.} \end{cases}$$

$$13. p_Y(y) = \begin{cases} \frac{2}{\pi \sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{其他.} \end{cases}$$

14. $W = 2I^2$ 的取值为 $162 < w < 242$.

$$F_W(w) = P(W \leq w) = P\left(I^2 \leq \frac{w}{2}\right) = \int_{-\sqrt{\frac{w}{2}}}^{\sqrt{\frac{w}{2}}} p_I(I) dI$$

$$= \int_9^{\sqrt{\frac{w}{2}}} \frac{1}{2} dx = \frac{1}{2} \left(\sqrt{\frac{w}{2}} - 9 \right),$$

所以
$$p_W(w) = \begin{cases} \frac{1}{4\sqrt{2w}}, & 162 < w < 242, \\ 0, & \text{其他.} \end{cases}$$

疑难解析

1. 为什么要引入随机变量?

答 概率统计是从数量上来研究随机现象统计规律的,为了便于数学上的推导和计算,必须把随机事件数量化,引入随机变量之后,使我们有可能利用数学分析的方法来研究随机试验,研究和分析其结果的规律性.因此,随机变量是研究随机试验的重要而有效的工具.

2. 随机变量与普通函数有什么区别?

答 ①随机变量的取值具有随机性,它随试验结果的不同而取不同的值,试验之前只知道它可能取值的范围,而不能预知它取什么值.

②随机变量的取值具有统计规律性,由于试验结果的出现有一定的概率,因而随机变量取各个值也有一定概率.

③随机变量是定义在随机试验的所有可能结果的集合上的函数,普通函数定义在实数轴上.

3. 引入随机变量的分布函数有什么意义?

答 分布函数 $F(x) = P(X \leq x)$ 给出了随机变量 X 的取值不大于实数 x 的概率,而 X 在任意区间 $(x_1, x_2]$ 上的概率也可用分布函数表示,即 $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$. 因此,分布函数完整地描述了随机变量的统计规律性.

另一方面,分布函数是一个普通实值函数,这样就可以用高等数学中的理论和方法加以研究分析,认识问题.

4. 超几何分布、二项分布和 Poisson 分布之间有一种什么关系?

答 这三种分布是非常重要的离散型随机变量的概率分布.有时,它们的概率计算会十分烦琐.当试验次数 n 很大时,可以推导出这三种分布间有一种近似关系式:

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}.$$

第一个等式要求 N 很大, $\frac{n}{N}$ 很小, 取 $p = \frac{M}{N}$ 即成立.

第二个等式要求 n 很大, p 很小, 取 $\lambda = np$ 即成立. 在实际计算中, 当 $n \geq 10$, $p \leq 0.1$ 时即可使用近似公式, 而 Poisson 分布可通过查表计算.

5. 为什么说正态分布是概率论中最重要的分布?

答 首先, 正态分布有着极其广泛的实际背景, 例如测量的误差、炮弹的弹着点的分布、人体生理特征的数量指标(身高、体重等)、产品的数量指标(直径、长度、体积、重量等), 都服从或近似服从正态分布. 可以说, 一个变量如果受到大量微小的、相互独立的随机因素影响, 那么这个变量一般是一个正态随机变量.

其次, 有的分布(如二项分布、Poisson 分布)的极限分布是正态分布; 有的分布(如 χ^2 分布、 t 分布)可通过正态分布导出. 因此, 不论在实际中还是在理论中, 正态分布都是最重要的一种分布.

6. 连续型随机变量 X 的概率密度函数是否是连续函数? 是否唯一?

答 连续型随机变量 X 的概率密度函数 $p(x)$ 不一定是连续函数, 但至多有有限个不连续点.

概率密度函数 $p(x)$ 不是唯一的. 如果改变 $p(x)$ 在有限个点的函数值, 得到新的非负函数 $f(x)$, 则随机变量的分布函数 $F(x)$ 可表示为 $f(x)$ 在 $(-\infty, x]$ 上的积分, 因此 $f(x)$ 也是连续型随机变量 X 的概率密度函数.

7. 不同的随机变量的分布函数是否一定不同?

答 不一定.

例如, 进行射击试验, 令

$$X_1 = \begin{cases} 1, & \text{射中;} \\ -1, & \text{没中;} \end{cases} \text{ 或 } X_2 = \begin{cases} -1, & \text{射中;} \\ 1, & \text{没中;} \end{cases}$$

X_1 与 X_2 是两个不同的随机变量, 但是它们有相同的分布函数:

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

8. 离散型随机变量的函数是否一定是离散型随机变量? 连续型随机变量的函数是否一定是连续型随机变量?

答 若 X 是离散型随机变量, 它的取值是有限个或可列无限多个, 因此 $Y = f(X)$ 的取值也是有限个或可列无限多个, 所以 $Y = f(X)$ 是离散型随机变量, 如 2.3.2 节典型题分析中的例 2.3.11.

若 X 是连续型随机变量, 那么 $Y = f(X)$ 可能是连续型随机变量, 如 2.3.2 节典型题型分析中的例 2.3.12, 也可能是离散型随机变量, 如 2.3.2 节典型题型分析中的例 2.3.17.

9. 为什么分布函数 $F(x)$ 定义为右连续?

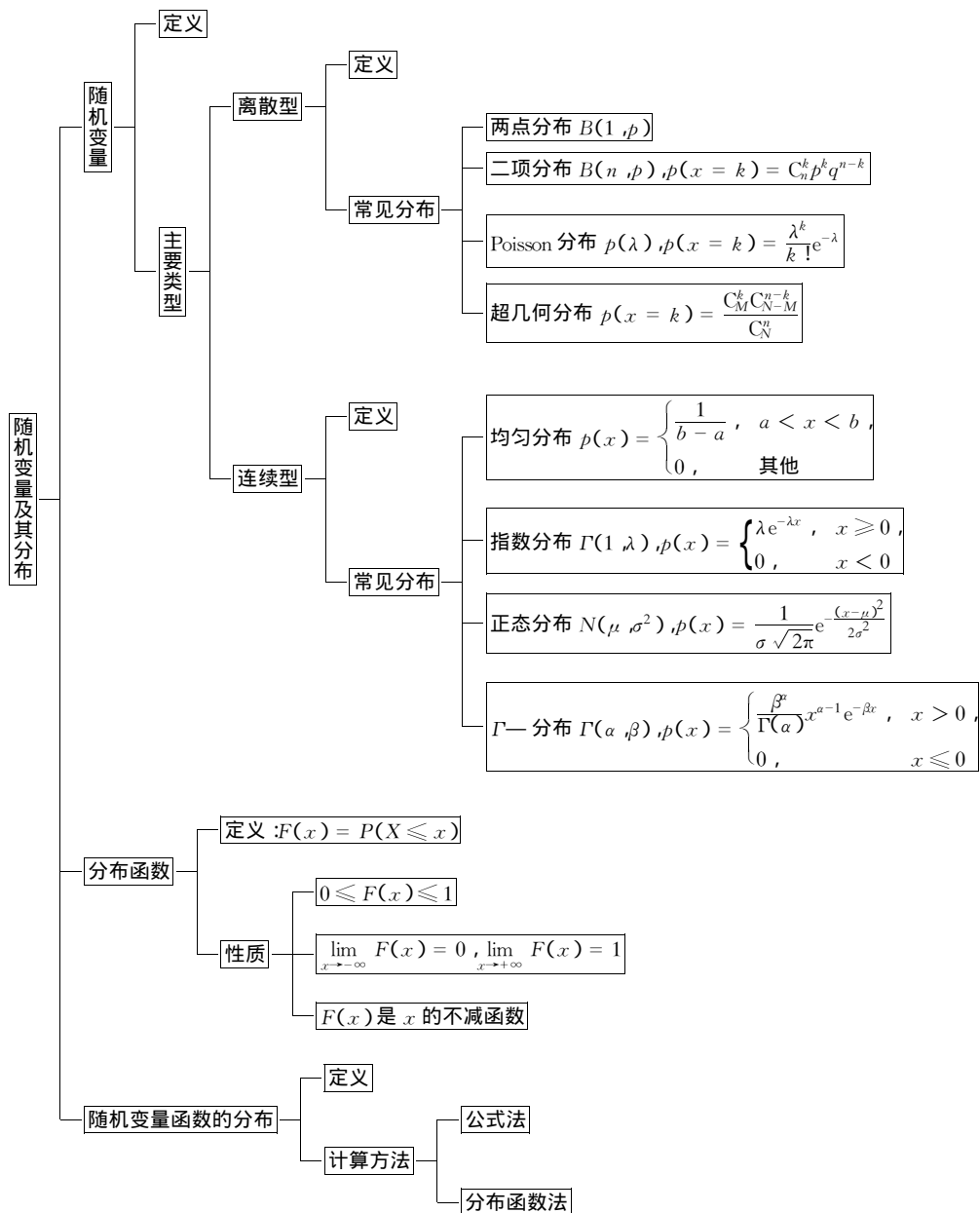
答 定义为左连续 $F(x) = P(X < x)$ 或者右连续 $F(x) = P(X \leq x)$, 只是一种习惯. 目

前,俄罗斯和东欧国家一般定义左连续,西欧和美国一般定义右连续,我国的大多书籍采用右连续.对于连续型随机变量而言,因为一点上的概率等于零,定义左连续和右连续没有什么区别.对于离散型随机变量,不同之处就在于 $X = x$ 点的概率是否计算在内.

本章常考题型归纳

考点	常考题型	例题
离散型随机变量的概率分布的概念、性质与计算 连续型随机变量的概率密度的概念、性质与计算 随机变量分布函数的概念、性质与计算 随机变量函数的分布 常见的离散型与连续型随机变量的概率分布	有关概率分布性质的命题	例 2.1.1~例 2.1.2
	计算离散型随机变量的概率分布	例 2.1.3~例 2.1.5
	利用分布列计算概率	例 2.1.6~例 2.1.12
	有关连续型随机变量概率密度性质的命题	例 2.2.1~例 2.2.3
	利用概率密度计算概率	例 2.2.4~例 2.2.9
	有关分布函数性质的命题	例 2.3.1
	已知概率分布(或概率密度)求分布函数	例 2.3.2~例 2.3.3
	已知分布函数求概率分布(或概率密度)	例 2.3.4~例 2.3.5
	有关随机变量函数的概率分布的计算题	例 2.3.10~例 2.3.17
	有关正态分布的计算题	例 2.3.6~例 2.3.9

本章知识结构网络图



检测题(一)

一、选择题

1. 设函数 $p(x)$ 在区间 $[a, b]$ 上等于 $\sin x$, 而在此区间外等于 0, 若 $p(x)$ 可以作为某连续型随机变量 X 的密度函数, 则区间 $[a, b]$ 为().

A. $\left[0, \frac{\pi}{2}\right]$

B. $[0, \pi]$

C. $\left[-\frac{\pi}{2}, 0\right]$

D. $\left[0, \frac{3}{2}\pi\right]$

2. 下列 $F(x)$ 中, 可以作为某随机变量 X 分布函数的是().

$$A. F(x) = \begin{cases} 0, & x \leq 0, \\ 0.3, & 0 < x \leq 1, \\ 0.2, & 1 < x \leq 2, \\ 1, & x > 2 \end{cases}$$

$$B. F(x) = \begin{cases} 0.5e^x, & x \leq 0, \\ 0.8, & 0 < x \leq 1, \\ 1, & x > 1 \end{cases}$$

$$C. F(x) = \begin{cases} 0, & x \leq 0, \\ 0.1x, & 0 < x \leq 5, \\ 0.4, & 5 < x \leq 6, \\ 1, & x > 6 \end{cases}$$

$$D. F(x) = \begin{cases} 0, & x \leq -\frac{\pi}{2}, \\ \sin x, & -\frac{\pi}{2} < x \leq 0, \\ 1, & x > 0 \end{cases}$$

3. 若随机变量 $X \sim N(\mu, \sigma^2)$ 则 $Y = aX + b$ 服从().

$$A. N(\mu, \sigma^2)$$

$$B. N(0, 1)$$

$$C. N\left(\frac{\mu}{a}, \frac{\mu^2}{b^2}\right)$$

$$D. N(a\mu + b, a^2\sigma^2)$$

4. 已知 X 的分布律为 $P(X=k) = \frac{1}{c} \cdot \frac{\lambda^k}{k!} (k=1, 2, \dots)$ 则 c 等于().

$$A. e^{-\lambda}$$

$$B. e^{\lambda}$$

$$C. e^{-\lambda-1}$$

$$D. e^{\lambda} - 1$$

5. 设连续型随机变量 X 的概率密度为 $p(x)$, 分布函数为 $F(x)$, 则下列选项中正确的是().

$$A. 0 \leq p(x) \leq 1$$

$$B. P(X=x) = F(x)$$

$$C. P(X \leq x) = F(x)$$

$$D. P(X=x) = p(x)$$

6. 已知随机变量 X 的概率密度为 $p(x) = \begin{cases} x, & 0 \leq x < 1, \\ 2-x, & 1 \leq x < 2, \\ 0, & \text{其他}, \end{cases}$ 则 $P(X \leq 1.5)$ 等于

().

$$A. 0.875$$

$$B. \int_0^{1.5} (2-x) dx$$

$$C. \int_0^{+\infty} p(x) dx$$

$$D. \int_{-\infty}^{1.5} (2-x) dx$$

7. 设 $P(X=k) = C \frac{\lambda^k e^{-\lambda}}{k!} (k=0, 1, \dots)$ 是随机变量 X 的概率分布, 则 λ 和 C 一定满足().

$$A. \lambda > 0$$

$$B. C < 0$$

$$C. C \cdot \lambda > 0$$

$$D. C > 0 \text{ 且 } \lambda > 0$$

8. 设 X 在 $[0, 1]$ 上服从均匀分布, $Y = 2X + 1$, 则下列选项正确的是().

$$A. Y \text{ 在 } [0, 1] \text{ 上服从均匀分布}$$

$$B. Y \text{ 在 } [1, 3] \text{ 上服从均匀分布}$$

$$C. P(0 \leq Y \leq 1) = 1$$

$$D. P(0 \leq Y < 1) = 0.5$$

9. 设随机变量 X 的分布函数为 $F(x)$, 在下列概率中可表示为 $F(a) - F(a-0)$ 的是().

$$A. P(X \leq a)$$

$$B. P(X > a)$$

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

$$D. P(X \geq a)$$

10. 若随机变量 $X \sim N(0, 1)$, 分布函数为 $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt (-\infty < x < +\infty)$, 且 $P(X > x) = a$, 则 $x =$ ().

$$A. \Phi^{-1}(a)$$

$$B. \Phi^{-1}\left(1 - \frac{a}{2}\right)$$

$$C. \Phi^{-1}(1-a)$$

$$D. \Phi^{-1}\left(\frac{a}{2}\right)$$

二、填空题

1. 已知离散型随机变量 X 只能取 $-1, 0, 1, 2$ 四个值, 相应概率为 $\frac{1}{2c}, \frac{3}{4c}, \frac{5}{8c}, \frac{7}{16c}$, 则 $c =$

_____. $P(X < 1 | X \neq 0) =$ _____.

2. 设一小时内进入图书馆的读者数服从 Poisson 分布. 已知一小时内无读者走进该图书馆的概率为 0.01, 则一小时内至少有一名读者进入图书馆的概率为 _____.

3. 设随机变量 X 的概率密度为 $p(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$ 以 Y 表示对 X 的三次独立重复

观察中事件“ $X \leq \frac{1}{2}$ ”出现的次数, 则 $P(Y=2) =$ _____.

4. 已知随机变量 $X \sim N(2, 3^2)$, 且 $P(X < c) = P(X \geq c)$, 则 $c =$ _____.

5. 设随机变量 X 服从 $p = \frac{1}{3}$ 的两点分布, $Y = 2X + 1$, 则 Y 的分布函数 $F_Y(y) =$ _____.

三、解答题

1. 设随机变量 X 的分布为

X	1	2	...	n	...
P	$\frac{1}{2}$	$\frac{1}{2^2}$	...	$\frac{1}{2^n}$	...

求 $Y = f(X) = \begin{cases} -1, & X \text{ 为奇数} \\ 1, & X \text{ 为偶数} \end{cases}$ 的分布律.

2. 设 10 件产品中有 7 件正品 3 件次品, 随机地从中抽取产品, 每次取 1 件, 直到取出正品为止, 则

(1) 若有放回地抽取, 求抽取次数的概率分布和分布函数;

(2) 若无放回地抽取, 求抽取次数的概率分布和分布函数.

3. 设随机变量 X 的概率密度为 $p_X(x) = \begin{cases} Ax, & 0 < x \leq 1, \\ B - x, & 1 \leq x \leq 2, \text{连续,} \\ 0, & \text{其他} \end{cases}$

求: (1) 常数 A, B ;

(2) 分布函数 $F(x)$;

(3) $P\left(\frac{1}{2} < X \leq \frac{3}{2}\right)$;

(4) $Y = 1 - \sqrt[3]{X}$ 的概率密度.

4. 若随机变量 X 服从参数为 2 的指数分布, 则 $Y = 1 - e^{-2X}$ 服从什么分布?

5. 设一页书上印刷错误数服从 Poisson 分布, 若某书有 500 页, 其中 100 页上没有印刷错误, 求该书至少有 2 个错误的页数约占百分之几?

检测题(一)参考答案

一、选择题

1. A 2. B 3. D 4. D 5. C 6. A 7. C 8. B 9. C 10. C

二、填空题

1. $\frac{37}{16}, \frac{8}{25}$

2. $1 - e^{-2 \ln 10} \approx 0.99$

$$3. \frac{9}{64}$$

$$4.2$$

$$5. F_Y(y) = \begin{cases} 0, & y < 0, \\ \frac{2}{3}, & 1 \leq y < 3, \\ 1, & y \geq 3 \end{cases}$$

三、解答题

$$1. \begin{array}{c|cc} Y & -1 & 1 \\ \hline P & \frac{2}{3} & \frac{1}{3} \end{array}.$$

$$2. (1) P(X=k) = \left(\frac{3}{10}\right)^{k-1} \frac{7}{10} (k=1, 2, \dots)$$

$$F(x) = 1 - \left(\frac{3}{10}\right)^{\lceil x \rceil}.$$

$$(2) \begin{array}{c|cccc} X & 1 & 2 & 3 & 4 \\ \hline P & \frac{7}{10} & \frac{7}{30} & \frac{7}{120} & \frac{1}{120} \end{array},$$

$$F(x) = \begin{cases} 0, & x < 1, \\ \frac{7}{10}, & 1 \leq x < 2, \\ \frac{28}{30}, & 2 \leq x < 3, \\ \frac{119}{120}, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

$$3. (1) A=1, B=2.$$

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

$$(3) P\left(\frac{1}{2} < X \leq \frac{3}{2}\right) = \frac{3}{4}.$$

$$(4) p_Y(y) = \begin{cases} 3(1-y)^5, & 0 \leq y < 1, \\ 3[2 - (1-y)^3][1-y]^2, & 1 - \sqrt[3]{2} \leq y < 0, \\ 0, & \text{其他.} \end{cases}$$

$$4. p_Y(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{其他,} \end{cases} Y \text{ 服从 } (0, 1) \text{ 上的均匀分布.}$$

$$5. P(X=0) = e^{-\lambda} = \frac{100}{500} \Rightarrow e^{-\lambda} = \frac{1}{5}, \lambda = \ln 5.$$

$$P(X=k)=\frac{1}{5}\frac{(\ln 5)^k}{k!},$$

$$P(X\geq 2)=1-\frac{1}{5}-\frac{1}{5}\frac{\ln 5}{1!}=0.4781.$$

检测题(二)

一、选择题

1. 设随机变量 X 的密度函数为 $p(x)$, 且 $p(-x)=p(x)$, $F(x)$ 是 X 的分布函数, 则对任意实数 a , 有().

A. $F(-a)=1-\int_0^a p(x)dx$

B. $F(-a)=\frac{1}{2}-\int_0^a p(x)dx$

C. $F(-a)=F(a)$

D. $F(-a)=2F(a)-1$

2. 设随机变量 X 与 Y 服从正态分布, $X\sim N(\mu, \sigma^2)$, $Y\sim N(\mu, \sigma^2)$. $p_1=P(X\leq \mu-4)$, $p_2=P(Y\geq \mu+5)$. 则().

A. 对任何实数 μ 均有 $p_1=p_2$

B. 对任何实数 μ 均有 $p_1<p_2$

C. 只对个别 μ 才有 $p_1=p_2$

D. 对任何实数 μ 均有 $p_1>p_2$

3. 在下述函数中, 可以作为某个随机变量的分布函数的是().

A. $F(x)=\frac{1}{1+x^2}$

B. $F(x)=\frac{1}{\pi}\arctan x+\frac{1}{2}$

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

D. $F(x)=\int_{-\infty}^x f(t)dt$ 其中 $\int_{-\infty}^{+\infty} f(t)dt=1$

4. 设离散型随机变量 X 的分布律为 $P(X=k)=b\lambda^k$ ($k=1, 2, \dots$) 且 $b>0$, 则 λ 为().

A. 大于 0 的任意实数

B. $\lambda=\frac{1}{b}$

C. $\lambda=\frac{1}{1+b}$

D. $\lambda=\frac{1}{b-1}$

5. 设 $F_1(x)$ 与 $F_2(x)$ 分别是随机变量 X_1 与 X_2 的分布函数, 为了使 $F(x)=aF_1(x)+bF_2(x)$ 是某一随机变量的分布函数, 在下面各组值中应取().

A. $a=\frac{3}{5}, b=-\frac{2}{5}$

B. $a=\frac{2}{3}, b=\frac{2}{3}$

C. $a=-\frac{1}{2}, b=\frac{3}{2}$

D. $a=\frac{1}{2}, b=-\frac{3}{2}$

6. 设随机变量 X 服从指数分布, 则对于随机变量 $Y=\min\{X, 2\}$ 的分布函数, 下面结论正确的是().

A. 连续函数

B. 至少有两个间断点

C. 阶梯函数

D. 恰好有一个间断点

7. 设随机变量 X 服从正态分布 $X\sim N(\mu, \sigma^2)$, 则随着 a 的增大, 概率 $p(|X-\mu|<a)$ 应().

A. 单调增大

B. 单调减少

C. 保持不变

D. 增减不定

8. 设随机变量 $X\sim N(1, 1)$, 概率密度为 $p(x)$ 则().

A. $P(X\leq 0)=P(X\geq 0)=0.5$

B. $p(x)=p(-x), x\in(-\infty, +\infty)$

C. $P(X\leq 1)=P(X\geq 1)=0.5$

D. $F(x)=1-F(-x), x\in(-\infty, +\infty)$

9. 设 $F(x)=\begin{cases} 0, & x\leq 0, \\ \frac{x}{2}, & 0<x\leq 1, \\ 1, & x>1, \end{cases}$ 则().

A. $F(x)$ 是随机变量 X 的分布函数B. $F(x)$ 不是分布函数C. $F(x)$ 是离散型分布函数D. $F(x)$ 是连续型分布函数10. 若随机变量 X 的密度函数为 $p_X(x) = \frac{1}{\pi(1+x^2)}$, 则 $Y=2X$ 的密度函数是()A. $\frac{1}{\pi(1+4y^2)}$ B. $\frac{1}{\pi(4+y^2)}$ C. $\frac{1}{\pi(1+y^2)}$ D. $\frac{1}{\pi} \arctan y$

二、填空题

1. 设随机变量 X 的分布函数为 $F(x) = \begin{cases} 0, & x < -1, \\ 0.4, & -1 \leq x < 1, \\ 0.8, & 1 \leq x < 3, \\ 1, & x \geq 3, \end{cases}$ 则 X 的概率分布为_____.2. 设随机变量 $X \sim B(2, p)$, $Y \sim B(3, p)$ 若 $P(X \geq 1) = \frac{5}{9}$ 则 $p(Y \geq 1) =$ _____.3. 设随机变量 X 的分布函数为 $F(x) = \begin{cases} 0, & x < 0, \\ A \sin x, & 0 \leq x \leq \frac{\pi}{2}, \\ 1, & x > \frac{\pi}{2}, \end{cases}$ 则 $P\left(|X| < \frac{\pi}{6}\right) =$ _____.4. 若随机变量 Y 在 $(1, 6)$ 上服从均匀分布, 则方程 $x^2 + Yx + 1 = 0$ 有实根的概率是_____.5. 若随机变量 $X \sim N(2, \sigma^2)$, 且 $P(2 < X < 4) = 0.3$ 则 $P(X < 0) =$ _____.

三、解答题

1. 一汽车沿一街道行驶, 需要通过三个设有红绿信号灯的路口, 每个信号灯为红或绿与其他信号灯为红或绿相互独立, 且红绿两种信号显示的时间相等. 求该汽车首次遇到红灯前已通过的路口个数的分布.

2. 在电源电压不超过 200 V, 在 200~240 V 和超过 240 V 的三种情形下, 某种电子元件损坏的概率分别为 0.1, 0.001 和 0.2. 假设电源电压 X 服从正态分布 $X \sim N(220, 25^2)$. 求:

① 电子元件损坏的概率;

② 电子元件损坏时, 电源电压在 200~240 V 之间的概率.

附表:

x	0.4	0.6	0.8	1	1.2
$\Phi(x)$	0.655	0.726	0.788	0.841	0.885

3. 假设测量的随机误差 $X \sim N(0, 10^2)$. 求在 100 次独立重复测量中, 至少有三次测量误差的绝对值大于 19.6 的概率 α , 并利用 Poisson 分布求出 α 的近似值(保留两位小数).

附表:

λ	1	2	3	4	5	6	7
$e^{-\lambda}$	0.368	0.135	0.05	0.018	0.007	0.002	0.001

4. 某仪器装有三只独立工作的同型号电子元件, 其寿命都服从指数分布, 其密度为 $p(x)$

$$= \begin{cases} \frac{1}{600} e^{-\frac{x}{600}}, & x > 0 \\ 0, & x \leq 0 \end{cases} \text{ 求在仪器使用的最初 200 小时内, 至少有一只电子元件损坏的概率.}$$

5. 设随机变量 X 的概率密度为 $p_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ 求 $Y = e^X$ 的概率密度 $p_Y(y)$.

6. 设随机变量 X 的概率密度为 $p(x) = \frac{1}{\pi(1+x^2)}$ 求 $Y = 1 - \sqrt[3]{X}$ 的概率密度.

7. 已知随机变量 X 的分布为 $P(X=k) = \frac{2}{3^k} (k=1, 2, \dots)$. 求 $Y = 1 + (-1)^X$ 的分布.

8. 设随机变量 X 取值为 $0, \frac{\pi}{2}, \pi, \dots, \frac{n}{2}\pi, \dots$ 的概率分别为 $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \frac{1}{2^{n+1}}, \dots$. 求 $Y = \sin X$ 的分布.

检测题(二)参考答案

一、选择题

1. B 2. A 3. B 4. C 5. A 6. D 7. C 8. C 9. A 10. B

二、填空题

1.

X	-1	1	3
P	0.4	0.4	0.2

 2. $\frac{19}{27}$ 3. $\frac{1}{2}$ 4. $\frac{4}{5}$ 5. 0.2

三、解答题

1.

X	0	1	2	3
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

.

2. (1) $0.212 \times 0.1 + 0.576 \times 0.001 + 0.212 \times 0.2 = 0.064$ 2.

(2) $\frac{0.576 \times 0.001}{0.064} = 0.009$.

3. $P\{|X| > 19.6\} = 2 - 2\Phi(1.96) = 0.05$,

$\alpha = 1 - 0.95^{100} - 100 \times 0.95^{99} \times 0.05 - \frac{100 \times 99}{2} \times 0.95^{98} \times 0.05^2$,

$\alpha \approx 1 - e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2}\right) = 0.87 (\lambda = 100 \times 0.05 = 5)$.

4. $1 - (e^{-\frac{1}{3}})^3 = 1 - e^{-1}$.

5. $p_Y(y) = \begin{cases} 0, & y < 1, \\ \frac{1}{y^2}, & y \geq 1. \end{cases}$ 6. $p_Y(y) = \frac{3}{\pi} \frac{(1-y)^2}{1+(1-y)^6}$. 7.

Y	0	2
P	$\frac{3}{4}$	$\frac{1}{4}$

.

8.

Y	-1	0	1
P	$\frac{1}{15}$	$\frac{2}{3}$	$\frac{4}{15}$

.

第 3 章 随机变量的数字特征

3.1 离散型随机变量的期望

3.1.1 知识要点

1. 定义

设离散型随机变量 X 的概率分布为

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

如果级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛, 则称它为 X 的数学期望, 记作 $E(X)$, 即

$$E(X)=\sum_{k=1}^{\infty} x_k p_k = x_1 p_1 + x_2 p_2 + \dots + x_k p_k + \dots$$

2. 几个常见分布期望

① 两点分布 设 $X \sim B(1, p)$ 则 $E(X)=p$;

② 二项分布 设 $X \sim B(n, p)$ 则 $E(X)=np$;

③ Poisson 分布 设 $X \sim P(\lambda)$ 则 $E(X)=\lambda$;

④ 超几何分布 设 $X \sim H(n, M, N)$ 则 $E(X)=\frac{nM}{N}$.

3.1.2 典型题型解析

例 3.1.1 某篮球运动员投篮命中率是 p , 现连续投篮直到投中为止, 求投篮次数 X 的数学期望.

解 $P(X=k)=(1-p)^{k-1}p, k=1,2,\dots$

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k \cdot q^{k-1} p \quad (\text{其中 } q=1-p) \\ &= p \sum_{k=1}^{\infty} k q^{k-1} = \frac{p}{(1-q)^2} \\ &= \frac{1}{p}. \end{aligned}$$

注 在未给出概率分布的情况下求 $E(X)$, 先要确定 X 的概率分布, 本题的概率分布通常称为几何分布, 记为 $X \sim G(p)$. 在利用定义求 $E(X)$ 的过程中, 使用了求和公式 $\sum_{k=1}^{\infty} k q^{k-1} =$

$\frac{1}{(1-q)^2}$, 它是在等式 $\sum_{k=1}^{\infty} q^k = \frac{1}{1-q} - 1$ 的两边关于 q 求导得到的.

例 3.1.2 一袋中装有 3 个红球 5 个白球, 从中抽取 4 次, 每次抽一球, 以 X 记取到红球的次数, 求:

(1) 在放回抽样下, X 的数学期望;

(2) 在无放回抽样下, X 的数学期望.

解 (1) 在放回抽样下, X 的可能值为 0, 1, 2, 3, 4, 且

$$P(X=0) = \left(\frac{5}{8}\right)^4, P(X=1) = C_4^1 \left(\frac{3}{8}\right) \left(\frac{5}{8}\right)^3 = \frac{3}{2} \times \left(\frac{5}{8}\right)^3,$$

$$P(X=2) = C_4^2 \left(\frac{3}{8}\right)^2 \left(\frac{5}{8}\right)^2 = 6 \times \frac{15^2}{8^4}, P(X=3) = C_4^3 \left(\frac{3}{8}\right)^3 \left(\frac{5}{8}\right) = \frac{5}{2} \times \left(\frac{3}{8}\right)^3,$$

$$P(X=4) = \left(\frac{3}{8}\right)^4,$$

所以
$$E(X) = 0 \times \left(\frac{5}{8}\right)^4 + 1 \times \frac{3}{2} \times \left(\frac{5}{8}\right)^3 + 2 \times 6 \times \frac{15^2}{8^4} + 3 \times \frac{5}{2} \times \left(\frac{3}{8}\right)^3 + 4 \times \left(\frac{3}{8}\right)^4 = \frac{3}{2}.$$

(2) 在无放回抽样下, X 的可能值为 0, 1, 2, 3, 且

$$P(X=0) = C_5^4 C_8^4 = \frac{5}{70}, P(X=1) = C_3^1 C_5^3 C_8^4 = \frac{30}{70},$$

$$P(X=2) = C_3^2 C_5^2 C_8^4 = \frac{30}{70}, P(X=3) = C_3^3 C_5^1 C_8^4 = \frac{5}{70},$$

所以
$$E(X) = 0 \times \frac{5}{70} + 1 \times \frac{30}{70} + 2 \times \frac{30}{70} + 3 \times \frac{5}{70} = \frac{3}{2}.$$

同步练习 3.1

1. 设随机变量 X 的概率分布为

X	-2	0	$\frac{1}{2}$	1
p	0.3	0.1	0.4	0.2

求 $E(X)$.

2. 已知随机变量 X 的概率分布为

X	4	6	x_3
p	p_1	0.3	0.2

又 $E(X) = 8$, 求 p_1 及 x_3 .

3. 假定每个人生日在各个月份的机会是相同的, 求三个人中生日在第一季度的人数的平均值.

同步练习 3.1 参考答案

1. $E(X) = -0.2$. 2. $p_1 = 0.5$, $x_3 = 21$. 3. 0.75.

3.2 连续型随机变量的期望

3.2.1 知识要点

1. 定义

设连续型随机变量 X 的概率密度为 $p(x)$, 如果积分 $\int_{-\infty}^{+\infty} xp(x)dx$ 绝对收敛, 则称此积分值为 X 的数学期望, 记作 $E(X)$. 即

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

2. 几个常见分布期望

① 均匀分布 设 $X \sim U(a, b)$ 则 $E(X) = \frac{a+b}{2}$;

② 指数分布 设 $X \sim \Gamma(1, \lambda)$ 则 $E(X) = \frac{1}{\lambda}$;

③ 正态分布 设 $X \sim N(\mu, \sigma^2)$ 则 $E(X) = \mu$;

④ Γ 分布 设 $X \sim \Gamma(\alpha, \beta)$ 则 $E(X) = \frac{\alpha}{\beta}$.

3. 如果随机变量 X 概率密度函数的图形以直线 $x = c$ 为对称轴, 则有 $E(X) = c$. 如均匀分布、正态分布即是如此.

3.2.2 典型题型解析

例 3.2.1 已知随机变量 X 的概率密度函数为 $p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2+2x-1}$, 求 $E(X)$.

解 将密度函数 $p(x)$ 改写为

$$p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2+2x-1} = \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{2}}} e^{-\frac{(x-1)^2}{2 \cdot \frac{1}{2}}},$$

因此, 可以看出 $X \sim N\left(1, \frac{1}{2}\right)$, 从而 $E(X) = 1$.

注 正态分布是一种重要的分布, 要熟悉它的密度函数形式.

例 3.2.2 设 X 的分布函数为

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

求 $E(X)$.

解 由于 $F'(x) = p(x)$ 所以

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

于是
$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xp(x)dx \\ &= \int_{-\infty}^0 x \cdot \frac{1}{2}e^x dx + \int_1^{+\infty} x \cdot \frac{1}{2}e^{-(x-1)} dx \\ &= -\frac{1}{2} + 1 = \frac{1}{2}. \end{aligned}$$

例 3.2.3 已知某电子元件的寿命(单位: h) X 有概率密度

$$p(x) = \begin{cases} a^2 x e^{-ax}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

其中 $a > 0$ 为常数. 试求电子元件的平均寿命.

解
$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xp(x)dx = a^2 \int_0^{+\infty} x^2 e^{-ax} dx \\ &= \frac{1}{a} \int_0^{+\infty} (ax)^2 e^{-ax} d(ax) \\ &= \frac{1}{a} \Gamma(3) = \frac{2}{a}. \end{aligned}$$

注 形如 $\int_0^{+\infty} x^\alpha e^{-x} dx$ ($\alpha > 0$ 为常数) 的积分是概率统计中经常会遇到的. 当 α 取较小整数时, 分部积分不失为一种好方法. 而当 α 取较大整数或非整数时, 常常借助 Γ 函数简化计算.

例 3.2.4 设随机变量 X 的概率密度为

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

求 $E(X)$.

解 将绝对值记号去掉, 化为

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

于是
$$E(X) = \int_0^1 x \cdot x dx + \int_1^2 x(2 - x) dx = \frac{1}{3} + \frac{2}{3} = 1.$$

注 若概率密度含有绝对值, 一般先将绝对值记号去掉.

同步练习 3.2

1. 设随机变量 X 的概率密度为 $p(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < +\infty$, 求 $E(X)$.
2. 设随机变量 X 的分布函数为

$$F(x) = \begin{cases} 0, & x \leq 0, \\ \frac{x}{4}, & 0 < x \leq 4, \\ 1, & x > 4, \end{cases}$$

求 $E(X)$.

3. 已知随机变量 X 的概率密度为 $p(x) = kx^a$, $0 < x < 2$. 其中 $a > 0$, k 为待定常数, 且 $E(X) = \frac{8}{5}$, 试求 a, k .

同步练习 3.2 参考答案

$$1. E(X) = 0. \quad 2. E(X) = 2. \quad 3. a = 3, k = \frac{1}{4}.$$

3.3 数学期望的性质与随机变量函数的期望公式

3.3.1 知识要点

1. 期望的性质

- ① $E(c) = c$ (c 为常数);
- ② $E(aX) = aE(X)$ (a 为常数);
- ③ $E(X + b) = E(X) + b$ (b 为常数);
- ④ $E(aX + b) = aE(X) + b$ (a, b 均为常数).

2. 随机变量函数的期望公式

设 $f(x)$ 连续, $Y = f(X)$ 是随机变量 X 的函数, 则

$$(1) \text{ 若 } X \text{ 为离散型 } E(Y) = E[f(X)] = \sum_{k=1}^{\infty} f(x_k) p_k;$$

$$(2) \text{ 若 } X \text{ 为连续型 } E(Y) = E[f(X)] = \int_{-\infty}^{+\infty} f(x) p(x) dx.$$

3.3.2 典型题型解析

例 3.3.1 已知随机变量 X 的概率密度为

$$p(x) = \begin{cases} ax, & 0 < x < 2, \\ cx + b, & 2 \leq x \leq 4, \\ 0, & \text{其他}, \end{cases}$$

且 $E(X) = 2$, $p(1 < x < 3) = \frac{3}{4}$ 求:

(1) a, b, c 的值; (2) $E(Y) = E(e^x)$.

解 (1) 由 $1 = \int_{-\infty}^{+\infty} p(x) dx = \int_0^2 ax dx + \int_2^4 (cx + b) dx$ 得

$$2a + 6c + 2b = 1;$$

$$\text{由 } 2 = \int_0^2 xax dx + \int_2^4 x(cx + b)dx \text{ 得 } \frac{8}{3}a + \frac{56}{3}c + 6b = 2;$$

$$\text{由 } \frac{3}{4} = P\{1 < x < 3\} = \int_1^2 ax dx + \int_2^3 (cx + b)dx \text{ 得}$$

$$\frac{3}{2} + \frac{5}{2}c + b = \frac{3}{4}.$$

$$\text{解方程组 } \begin{cases} 2a + 6c + 2b = 1, \\ \frac{8}{3}a + \frac{56}{3}c + 6b = 2, \\ \frac{3}{2} + \frac{5}{2}c + b = \frac{3}{4}, \end{cases} \text{ 得 } \begin{cases} a = \frac{1}{4}, \\ b = 1, \\ c = -\frac{1}{4}. \end{cases}$$

$$(2) E(Y) = E(e^x) = \int_0^2 e^x \cdot \frac{x}{4} dx + \int_2^4 e^x \left(1 - \frac{x}{4}\right) dx = \frac{1}{4}(e^2 - 1)^2.$$

注 求随机变量函数的期望常常需结合概率密度的性质和随机变量的概率等内容.

例 3.3.2 如果在 1 500 件产品中有 100 件不合格品,从中任意抽取 150 件进行检查,求查得不合格品数的数学期望.

解 设 X 为查得的不合格品数. 令

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 个产品为不合格品, } i = 1, 2, \dots, 150, \\ 0, & \text{第 } i \text{ 个产品为合格品,} \end{cases}$$

$$\text{则 } X = X_1 + X_2 + \dots + X_{150}.$$

$$\text{又因为 } \begin{array}{c|cc} X_i & 0 & 1 \\ \hline p & \frac{14}{15} & \frac{1}{15} \end{array} \text{ 所以 } E(X_i) = \frac{1}{15}, i = 1, 2, \dots, 150.$$

$$\text{因此 } E(X) = \sum_{i=1}^{150} E(X_i) = 150 \times \frac{1}{15} = 10.$$

注 本例方法经常用来求较复杂的但能分解为若干较简单的同分布的随机变量和的随机变量的数字特征.

例 3.3.3 游客乘电梯从底层到电视塔顶层观光,电梯于每个整点的 5 分钟、25 分钟和 55 分钟从底层起行,假设一游客在早上 8 点的第 X 分钟到达底层电梯处,且 X 在 $[0, 60]$ 上服从均匀分布,求该游客等候时间的数学期望.

解 已知 X 在 $[0, 60]$ 上服从均匀分布,其概率密度为

$$p(x) = \begin{cases} \frac{1}{60}, & 0 \leq x \leq 60, \\ 0, & \text{其他.} \end{cases}$$

设 Y 为游客等候电梯的时间(单位:分),则

$$Y = f(X) = \begin{cases} 5 - X, & 0 < X \leq 5, \\ 25 - X, & 5 < X \leq 25, \\ 55 - X, & 25 < X \leq 55, \\ 60 - X + 5, & 55 < X \leq 60. \end{cases}$$

$$\begin{aligned}
 \text{因此 } E(Y) &= E[f(X)] = \int_{-\infty}^{+\infty} f(x)p(x)dx = \frac{1}{60} \int_0^6 f(x)dx \\
 &= \frac{1}{60} \int_0^5 (5-x)dx + \frac{1}{60} \int_5^{25} (25-x)dx + \frac{1}{60} \int_{25}^{55} (55-x)dx + \frac{1}{60} \int_{55}^{60} (65-x)dx \\
 &= \frac{1}{60} (12.5 + 200 + 450 + 37.5) \\
 &= 11.67 (\text{分钟}).
 \end{aligned}$$

注 首先要找出 X 与 Y 的关系, 这是一个难点; 其次要注意如果游客是在第 55 分至 60 分之间到达, 则只能乘 9 点 5 分的电梯, 即等候时间应为 $60 - X + 5$ 分.

例 3.3.4 某厂生产的一种设备的使用寿命 X 的概率密度为

$$p(x) = \begin{cases} \frac{1}{4} e^{-\frac{x}{4}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

工厂规定: 已售设备在一年内损坏予以调换. 若工厂出售一台设备盈利 100 元, 调换一台设备厂方损失 300 元, 求厂方出售一台设备盈利值的数学期望.

解 厂方出售一台设备净盈利 Y 只取两个值 100 元和 -200 元, 于是

$$P(Y=100) = P(X \geq 1) = \int_1^{+\infty} \frac{1}{4} e^{-\frac{x}{4}} dx = e^{-\frac{1}{4}},$$

$$P(Y=-200) = P(X < 1) = 1 - e^{-\frac{1}{4}}.$$

$$\begin{aligned}
 \text{所以 } E(Y) &= 100 \times e^{-\frac{1}{4}} + (-200) \times (1 - e^{-\frac{1}{4}}) \\
 &= 300e^{-\frac{1}{4}} - 200 = 33.64 (\text{元}).
 \end{aligned}$$

注 解答本题首先要找到 X 和 Y 的关系, 值得注意的是 X 是连续型随机变量, 而 Y 是离散型的随机变量.

同步练习 3.3

1. 设 X 在 $[-\frac{1}{2}, \frac{1}{2}]$ 上服从均匀分布

$$y = f(x) = \begin{cases} \ln x, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

求随机变量 $Y = f(X)$ 的数学期望.

2. 已知离散型随机变量 X 服从参数为 2 的 Poisson 分布, 求随机变量 $Y = 3X - 2$ 的数学期望 $E(Y)$.

3. 现有 10 张奖券, 其中 8 张为 2 元的, 2 张为 5 元的, 今从中随机无放回地抽取 3 张, 求得奖金的数学期望.

4. 某机场的送客车一次载 20 名旅客自机场开出, 沿途有 10 个停车点, 若到达停车点无人下车则车不停下. 设每名旅客在各个停车点下车是等可能的, 求送客车停车次数的数学期望.

同步练习 3.3 参考答案

$$1. -\frac{1}{2}(\ln 2 + 1). \quad 2. 4. \quad 3. 7.8. \quad 4. 8.784.$$

3.4 方差及其性质

3.4.1 知识要点

1. 定义

① 设 X 是一个随机变量, 如果 $E\{[X - E(X)]^2\}$ 存在, 则称其为随机变量 X 的方差, 记作 $D(X)$, 即 $D(X) = E\{[X - E(X)]^2\}$.

若 X 为离散型 $D(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k$;

若 X 为连续型 $D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 p(x) dx$;

重要公式 $D(X) = E(X^2) - E^2(X)$.

② 设 X 为随机变量, 如果 $E(X^k) (k=1, 2, \dots)$ 存在, 则称其为 X 的 k 阶原点矩, 记为 v_k , 即 $v_k = E(X^k) (k=1, 2, \dots)$.

如果 $E\{[X - E(X)]^k\} (k=1, 2, \dots)$ 存在, 则称其为 X 的 k 阶中心矩, 记为 μ_k , 即

$$\mu_k = E\{[X - E(X)]^k\} (k=1, 2, \dots).$$

显然, $v_1 = E(X)$, $\mu_2 = D(X)$.

2. 几个常见分布的方差

① 两点分布 设 $X \sim B(1, p)$ 则 $D(X) = pq$;

② 二项分布 设 $X \sim B(n, p)$ 则 $D(X) = npq$;

③ Poisson 分布 设 $X \sim P(\lambda)$ 则 $D(X) = \lambda$;

④ 均匀分布 设 $X \sim U(a, b)$ 则 $D(X) = \frac{1}{12}(b-a)^2$;

⑤ 指数分布 设 $X \sim \Gamma(1, \lambda)$ 则 $D(X) = \frac{1}{\lambda^2}$;

⑥ 正态分布 设 $X \sim N(\mu, \sigma^2)$ 则 $D(X) = \sigma^2$;

⑦ Γ 分布 设 $X \sim \Gamma(\alpha, \beta)$ 则 $D(X) = \frac{\alpha}{\beta^2}$.

3. 方差的性质

设 X 为随机变量, a, b, c 均为常数, 则有

① $D(c) = 0$;

② $D(aX) = a^2 D(X)$;

③ $D(X+b) = D(X)$;

④ $D(aX+b) = a^2 D(X)$.

4. Chebyshev 不等式

设 X 为随机变量, $E(X) = \mu$, $D(X) = \sigma^2$, 则对任意的正数 ϵ 均有

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

称上式为 Chebyshev 不等式.

3.4.2 典型题型解析

例 3.4.1 已知随机变量 X 服从二项分布, 且 $E(X)=2.4$, $D(X)=1.68$, 求二项分布的参数 n, p .

解 因为 $X \sim B(n, p)$, 故 $E(X)=np$, $D(X)=np(1-p)$, 从而

$$np=2.4, np(1-p)=1.68,$$

所以 $p=0.3, n=8$.

注 要熟练掌握常见分布的数字特征, 从而可以确定分布中的参数的取值.

例 3.4.2 设随机变量 X 满足

$$E[(X-1)^2]=10, E[(X-2)^2]=6.$$

求 $E(X)$ 和 $D(X)$.

$$\text{解 } E[(X-1)^2]=E(X^2)-2E(X)+1=10,$$

$$\text{则 } E(X^2)-2E(X)=9,$$

$$E[(X-2)^2]=E(X^2)-4E(X)+4=6,$$

$$\text{则 } E(X^2)-4E(X)=2.$$

解联立方程组得 $E(X)=\frac{7}{2}$, $E(X^2)=16$. 于是

$$D(X)=E(X^2)-[E(X)]^2=16-\frac{49}{4}=\frac{15}{4}.$$

注 求随机变量的期望或方差并不一定要用公式求, 有时利用期望或方差的性质以及期望和方差的关系求更为简单.

例 3.4.3 设随机变量 X 的概率密度为

$$p(x)=\begin{cases} c(1-x^2)^\alpha, & -1 < x < 1 (\alpha < 0), \\ 0, & \text{其他.} \end{cases}$$

求 $E(X)$ 和 $D(X)$.

解 因为 $p(x)$ 是偶函数, 所以

$$E(X)=\int_{-\infty}^{+\infty} xp(x)dx=\int_{-1}^1 cx(1-x^2)^\alpha dx=0.$$

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

$$=\int_{-1}^1 cx^2(1-x^2)^\alpha dx$$

$$=-\frac{c}{2(\alpha+1)}\int_{-1}^1 x d(1-x^2)^{\alpha+1}$$

$$=-\frac{c}{2(\alpha+1)}x(1-x^2)^{\alpha+1}\Big|_{-1}^1 + \frac{c}{2(\alpha+1)}\int_{-1}^1 (1-x^2)^{\alpha+1} dx$$

$$=\frac{c}{2(\alpha+1)}\int_{-1}^1 (1-x^2)(1-x^2)^\alpha dx$$

$$= \frac{c}{2(\alpha+1)} \int_{-1}^1 (1-x^2)^\alpha dx - \frac{c}{2(\alpha+1)} \int_{-1}^1 x^2 (1-x^2)^\alpha dx,$$

而
$$c \int_{-1}^1 (1-x^2)^\alpha dx = \int_{-\infty}^{+\infty} p(x) dx = 1,$$

$$c \int_{-1}^1 x^2 (1-x^2)^\alpha dx = \int_{-\infty}^{+\infty} x^2 p(x) dx = D(X),$$

因此,有
$$D(X) = \frac{1}{2(\alpha+1)} - \frac{1}{2(\alpha+1)} D(X),$$

所以
$$D(X) = \frac{1}{2\alpha+3}.$$

注 本题利用了等式 $\int_{-\infty}^{+\infty} p(x) dx = 1$ 及 $\int_{-\infty}^{+\infty} x^2 p(x) dx = D(X)$, 不仅使计算变得简单, 而且也很巧妙.

例 3.4.4 甲、乙两家灯泡厂生产的灯泡的寿命 X 和 Y 的分布律为

X	900	1 000	1 100
p	0.1	0.8	0.1

,

Y	950	1 000	1 050
p	0.3	0.4	0.3

问哪家厂生产的灯泡质量好?

解
$$E(X) = 900 \times 0.1 + 1\,000 \times 0.8 + 1\,100 \times 0.1 = 1\,000,$$

$$E(Y) = 950 \times 0.3 + 1\,000 \times 0.4 + 1\,050 \times 0.3 = 1\,000.$$

由于 $E(X) = E(Y)$, 即平均寿命相同, 故需考察其性能的稳定性, 也就是必须比较方差的大小.

$$\begin{aligned} D(X) &= \sum_{k=1}^3 [x_k - E(X)]^2 p_k \\ &= (900 - 1\,000)^2 \times 0.1 + (1\,000 - 1\,000)^2 \times 0.8 + (1\,100 - 1\,000)^2 \times 0.1 \\ &= 2\,000, \end{aligned}$$

$$\begin{aligned} D(Y) &= \sum_{k=1}^3 [y_k - E(Y)]^2 p_k \\ &= (950 - 1\,000)^2 \times 0.3 + (1\,000 - 1\,000)^2 \times 0.4 + (1\,050 - 1\,000)^2 \times 0.3 \\ &= 1\,500. \end{aligned}$$

由此知 $D(Y) < D(X)$, 故乙厂生产的灯泡比较好.

注 方差是考察随机变量取值分散程度的一个数量指标, 方差越小说明随机变量的取值越集中在均值附近, 方差越大, 说明随机变量取值离均值越远.

例 3.4.5 设随机变量 X 服从指数分布, 其概率密度为

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

求 $D(\sqrt{x})$.

解 因为
$$E(\sqrt{x}) = \int_{-\infty}^{+\infty} \sqrt{x} p(x) dx = \int_0^{+\infty} \lambda x^{\frac{1}{2}} e^{-\lambda x} dx$$

$$= \frac{1}{\sqrt{\lambda}} \int_0^{+\infty} (\lambda x)^{\frac{1}{2}} e^{-\lambda x} d(\lambda x)$$

$$= \frac{1}{\sqrt{\lambda}} \Gamma\left(1 + \frac{1}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)}{2\sqrt{\lambda}} = \frac{\sqrt{\pi}}{2\sqrt{\lambda}}.$$

又由题设 $E(X) = \frac{1}{\lambda}$,

$$\text{故 } D(\sqrt{X}) = E(X) - [E(\sqrt{X})]^2 = \frac{1}{\lambda} - \frac{\pi}{4\lambda} = \frac{4-\pi}{4\lambda}.$$

同步练习 3.4

1. 已知 $E(X)=3, D(X)=5$, 试求 $E(X+2)^2$.
2. 一零件的横截面是圆, 对截面的直径进行测量, 设其直径 X 服从 $[0, 3]$ 上的均匀分布, 求横截面积 Y 的数学期望与方差.
3. 设随机变量 λ 服从参数为 λ 的 Poisson 分布, 若 $P(X=1)=P(X=2)$, 求 $E(X)$ 和 $D(X)$.
4. 设连续型随机变量 X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ kx + b, & 0 \leq x \leq \pi, \\ 1, & x > \pi. \end{cases}$$
 - ① 试确定常数 k, b 的值;
 - ② 求 $E(X), D(X)$;
 - ③ 设 $Y = \sin X$, 求 $E(Y)$.
5. 设随机变量 X 存在数学期望 $\mu = E(X)$, 方差 $\sigma^2 = D(X)$. 试证: 在线性变换 $X^* = \frac{X-\mu}{\sigma}$ 下, $E(X^*)=0, D(X^*)=1$.

同步练习 3.4 参考答案

1. 30. 2. $E(Y) = \frac{3}{4}\pi, D(Y) = \frac{9}{20}\pi$.
3. $E(X)=2, D(X)=2$.
4. $k = \frac{1}{\pi}, b=0, E(X) = \frac{\pi}{2}, D(X) = \frac{\pi^2}{12}, E(Y) = \frac{2}{\pi}$.

疑难解析

1. 随机变量的数字特征在概率论中有什么实际意义?

答 要全面掌握一个随机变量的统计规律性, 必须了解这个随机变量的分布函数. 但是, 在实际问题中要求得一个随机变量的分布函数是困难的, 而且也往往是不必要的. 通常只需要了解随机变量的某几个数量指标就可以了, 这就是概率论中的随机变量的数字特征. 一来, 随机变量的数字特征比较简单易求(在实际问题中可以通过样本进行估计); 二来, 已经足够满足解决实际问题的需要, 并且刻画了随机变量的某些特征.

2. 在数学期望定义中为什么要求级数 $\sum_{k=1}^{\infty} x_k p_k$ 和积分 $\int_{-\infty}^{+\infty} f(x) dx$ 绝对收敛?

答 这首先是因为,随机变量的取值是随机的,不一定是按顺序的,因此在求和的时候,可能要改变项的次序.其次是,由高等数学知,当级数绝对收敛时,级数不因改变项的次序而改变和,因而期望值唯一存在.积分是一个积分和式(也可以视作一个级数)的极限,因而也要求绝对收敛.

3. 随机变量的分布与数字特征有何关系?

答 随机变量的分布完全确定数字特征,反之则不然.例如: X, Y 的概率密度分别为

$$p_X(x) = ce^{-x^a \cos a\pi}, \quad x > 0;$$

$$p_Y(x) = c[1 + \sin(x^a \cos a\pi)]e^{-x^a \cos a\pi}$$

其中 $x > 0, 0 < a < \frac{1}{2}, c = a(\cos a\pi)^{\frac{1}{a}} \Gamma\left(\frac{1}{a}\right)$.

通过计算可知 $E(X) = E(Y)$, 但 $p_X(x) \neq p_Y(y)$.

4. 为什么不用 $E[X - E(X)]$ 或 $E[|X - E(X)|]$ 代替方差作为衡量 X 取值偏离程度的一个尺度?

答 因为 $X - E(X)$ 虽然反映了随机变量 X 的取值与平均值 $E(X)$ 的偏差,但偏差可能是负的,因而在求和时易于正负抵消,故 $E[X - E(X)]$ 的值不能正确反映 X 取值与平均值的偏离程度.

对 $E[|X - E(X)|]$ 来说,虽然不再存在正负值相互抵消的问题,但是,由于要取绝对值,给计算制造了一定的麻烦.所以, $E[X - E(X)]$ 和 $E[|X - E(X)|]$ 都不适合作为衡量随机变量 X 取值偏离程度的一个尺度.

5. 随机变量 X 的数学期望 $E(X)$ 与方差 $D(X)$ 之间有什么联系?

答 (1) 若 $E(X)$ 不存在,则 $D(X)$ 一定不存在.

这是因为,由公式 $D(X) = E(X^2) - [E(X)]^2$, 所以 $E(X)$ 不存在,必有 $D(X)$ 不存在.例如, X 服从柯西分布,即

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < +\infty.$$

由于 $\int_{-A}^A |x| \frac{dx}{\pi(1+x^2)} = \frac{2}{2\pi} \int_0^A \frac{d(1+x^2)}{1+x^2} = \frac{1}{\pi} \ln(1+x^2) \Big|_0^A \rightarrow \infty$ (当 $A \rightarrow +\infty$),

故 $E(X)$ 不存在,从而 $D(X)$ 也不存在.

(2) 若 $E(X)$ 存在, $D(X)$ 不一定存在.

(3) 若 $D(X)$ 存在,则对任意常数 C , 有 $D(X) \leq E(X - C)^2$, 当且仅当 $C = E(X)$ 时等号成立.

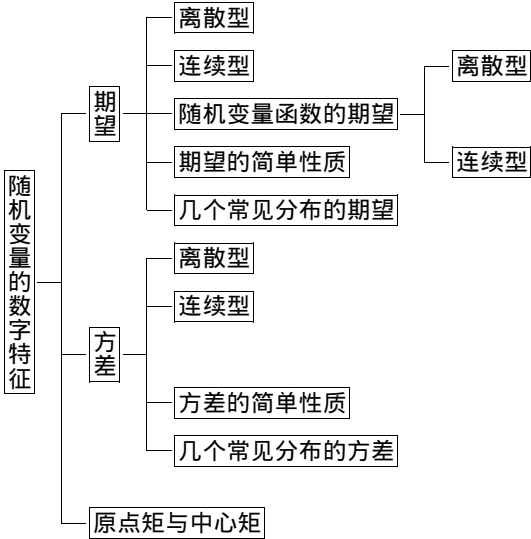
6. 随机变量的数字特征,就是指随机变量的期望与方差吗?

答 不对.所谓随机变量的数字特征,是刻画随机变量(或它的分布)的某些特征的数值.随机变量的数字特征主要有:数字期望 $E(X)$; 方差 $D(X)$; 矩等.

本章常考题型归纳

考点	常见题型	例题
由概率分布求期望、方差	已知离散型的概率分布求 $E(X)$ 和 $D(X)$	例 3.4.4
	已知连续型的概率密度求 $E(X)$ 和 $D(X)$	例 3.2.1 例 3.2.3 例 3.2.4 例 3.4.3
	已知分布函数求 $E(X)$ 和 $D(X)$	例 3.2.2
先求分布再求期望、方差	分布未知需建立概率分布求 $E(X)$ 和 $D(X)$	例 3.1.1 例 3.1.2 例 3.3.2 例 3.3.4
	分布未知需建立随机变量函数求 $E(X)$ 和 $D(X)$	例 3.3.3
期望方差性质	利用 $E(X)$ 和 $D(X)$ 的性质求参数	例 3.4.1 例 3.4.2
	已知随机变量函数求数字特征	例 3.3.1 例 3.4.5

本章知识结构网络



检测题(一)

一、选择题

1. 已知随机变量 X 服从二项分布 ,且 $E(X)=2.4$, $D(X)=1.44$,则二项分布的参数 n , p 的值为().
- A. $n=4$, $p=0.6$ B. $n=6$, $p=0.4$ C. $n=8$, $p=0.3$ D. $n=24$, $p=0.1$
2. 设随机变量 X 服从参数为 1 的指数分布 ,则随机变量 $Y=X+e^{-2X}$ 的数学期望 $E(Y)$ ().
- A. $\frac{3}{2}$ B. 5 C. $\frac{3}{4}$ D. $\frac{4}{3}$
3. 设 X 是一随机变量 , $E(X)=\mu$, $D(X)=\sigma^2$ (μ , $\sigma>0$ 为常数) ,则对任意常数 C ,也有

().

A. $E(X-C)^2 = E(X)^2 - C^2$

B. $E(X-C)^2 = E(X-\mu)^2$

C. $E(X-C)^2 < E(X-\mu)^2$

D. $E(X-C)^2 \geq E(X-\mu)^2$

4. 设随机变量 X 的期望 $E(X)$ 存在, 且 $E(X)=a$, $E(X^2)=b$, c 为一常数, 则 $D(cX)=$

().

A. $c(a-b^2)$

B. $c(b-a)^2$

C. $c^2(b-a^2)$

D. $c^2(a-b^2)$

二、填空题

1. 已知随机变量 X 的概率分布为

X	0	1	2	3
p	0.4	0.3	0.2	0.1

则 $E(X)=$ _____.2. 设随机变量 X 的概率密度 $p(x)=\begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{其他} \end{cases}$, 则 $D(X)=$ _____.3. 若随机变量 X 服从指数分布, 且 $D(X)=0.25$, 则 X 的概率密度为 _____.4. 设随机变量 X 服从参数为 λ 的 Poisson 分布, 且 $E(X-1)(X-2)=1$, 则 $\lambda=$ _____.

三、解答题

1. 盒中有 5 个球, 其中有 3 个白球, 2 个黑球, 从中任取 2 个球, 求白球数 X 的数学期望和方差.2. 设连续型随机变量 X 的分布函数为

$$F(x)=\begin{cases} 0, & x < -1, \\ a + b \arcsin x, & -1 \leq x < 1, \\ 1, & x \geq 1, \end{cases}$$

确定常数 a, b , 并求 $E(X)$ 和 $D(X)$.3. 设随机变量 X 在 $[-\frac{1}{2}, \frac{1}{2}]$ 上服从均匀分布, 又 $Y = \sin \pi X$, 试求随机变量 Y 的数学期望和方差.

4. 某银行开展定期定额有奖储蓄, 定期一年, 定额 60 元, 按规定 10 000 个户头中, 头等奖一个, 奖金 500 元, 二等奖 10 个, 各奖 100 元, 三等奖 100 个, 各奖 10 元, 四等奖 1 000 个, 各奖 2 元, 某人买了 5 个户头, 他期望得奖多少元?

5. 假设国际市场上每年对我国某种出口商品的需求量是随机变量 X (单位: t), 它服从于 $[2\ 000\ 4\ 000]$ 上的均匀分布. 设每售出这种商品 1 t, 可为国家挣得 3 万元, 但假如销售不出而屯积于库, 则每吨浪费保养费 1 万元, 问题是要组织多少吨货源, 才能使国家的收益最大?6. 某种稀有昆虫的寿命 X (约小时计) 具有概率密度

$$p(x)=\begin{cases} \frac{3}{4}(2x-x^2), & 0 < x < 2, \\ 0, & \text{其他} \end{cases}$$

试求 $E(X)$, $D(X)$.

检测题(一)参考答案

一、选择题

1. B 2. D 3. D 4. C

二、填空题

1. 1 2. $\frac{1}{18}$ 3. $p(x) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ 4. $\lambda = 1$

三、解答题

1. $E(X) = 1.2$, $D(X) = 0.46$.2. $a = \frac{1}{2}$, $b = \frac{1}{\pi}$, $E(X) = 0$, $D(X) = \frac{1}{2}$.3. $E(X) = 0$, $D(X) = \frac{1}{2}$.

4. 0.45 元.

5. 3 500 t.

6. $E(X) = 1$, $D(X) = 0.2$.

检测题(二)

一、选择题

1. 设随机变量 X 的分布函数为 $F(x) = \begin{cases} 0, & x < 0 \\ x^3, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$ 则 $E(X) = (\quad)$.A. $\int_0^{+\infty} x^4 dx$ B. $\int_0^1 3x^3 dx$ C. $\int_0^1 x^4 dx + \int_1^{+\infty} x dx$ D. $\int_0^{+\infty} 3x^3 dx$ 2. 设随机变量 Y 的函数 $Y = aX + b$ (a, b 为常数) 且 $E(X), D(X)$ 均存在, 则必有 ().A. $E(Y) = aE(X)$ B. $D(Y) = aD(X)$ C. $E(Y) = aE(X) + b$ D. $D(Y) = aD(X) + b$ 3. 如果随机变量 X 服从 () 上的均匀分布, 则 $E(X) = 3, D(X) = \frac{4}{3}$.A. $[0, 6]$ B. $[1, 5]$ C. $[2, 4]$ D. $[-3, 3]$ 4. 设随机变量 $X \sim N(0, 1), Y = 2X + 1$ 则 $Y \sim (\quad)$.A. $N(1, 4)$ B. $N(0, 1)$ C. $N(1, 1)$ D. $N(1, 2)$ 5. 设随机变量 X 的期望 $E(X)$, 方差 $D(X)$ 及 $E(X^2)$ 都存在, 则一定有 ().A. $E(X) \geq 0$ B. $D(X) \geq 0$ C. $E(X^2) \geq [E(X)]^2$ D. $E(X^2) \geq E(X)$ 6. 设随机变量 X 的密度函数为 $p(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$ 则 $P\{|X - E(X)| \geq 2\sqrt{DX}\}$

等于 ().

A. $\frac{9-8\sqrt{2}}{9}$ B. $\frac{6+4\sqrt{2}}{9}$ C. $\frac{6-4\sqrt{2}}{9}$ D. $\frac{9+8\sqrt{2}}{9}$

二、填空题

1. 设随机变量 X 服从某一区间上的均匀分布, 且 $E(X)=3$, $D(X)=\frac{1}{3}$, 则 X 的概率密度为_____, $P(X=2)=$ _____, $P(1<X<3)=$ _____.

2. 若随机变量 X 服从均值为 2, 方差为 σ^2 的正态分布, 且 $P(2<X<4)=0.3$, 则 $P(X<0)=$ _____.

3. 设 X 是一个随机变量, 其概率密度为 $p(x)=\begin{cases} 1+x, & -1\leq x<0, \\ 1-x, & 0\leq x\leq 1, \\ 0, & \text{其他,} \end{cases}$

则方差 $D(X)=$ _____.

4. 已知离散型随机变量 X 服从参数为 2 的 Poisson 分布, 即 $P(X=k)=\frac{2^k e^{-2}}{k!}$, 则随机变量 $Y=3X-2$ 的数学期望 $E(Y)=$ _____.

三、计算题

1. 设随机变量 X 的概率密度为 $p(x)=\frac{1}{2}e^{-|x|}$, $-\infty<x<+\infty$, 求 X 的数学期望 $E(X)$ 和方差 $D(X)$.

2. 将 4 个球随机地放入 4 个盒子里去, 每个球放入各个盒子是等可能的, 求没有球的盒子数 X 的数学期望.

3. 某公共汽车站每隔 5 min 有一辆汽车停站, 乘客在任意时刻到达汽车站, 求候车时间的数学期望(设汽车到站后乘客都能上车).

4. 已知随机变量 X 的概率密度形式为

$$p(x)=\begin{cases} ax, & 0<x<2, \\ cx+b, & 2\leq x\leq 4, \\ 0, & \text{其他,} \end{cases}$$

且 $E(X)=2$, $P(1<X<3)=\frac{3}{4}$.

求 (1) a, b, c 的值 (2) $E(Y)=E(e^X)$.

5. 设随机变量 X 满足 $E[(X-1)^2]=10$, $E[(X-2)^2]=6$, 求 $E(X)$ 与 $D(X)$.

6. 设随机变量 X 服从拉普拉斯分布, 其概率密度为

$$p(x)=\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}} \quad (\lambda>0),$$

求 $E(X)$ 和 $D(X)$.

检测题(二)参考答案

一、选择题

1. B 2. C 3. B 4. A 5. B 6. C.

二、填空题

1. $p(x)=\begin{cases} \frac{1}{2}, & 2\leq x\leq 4, \\ 0, & \text{其他,} \end{cases}$

$$2. 0.2 \quad 3. \frac{1}{6} \quad 4. 4$$

三、计算题

$$1. E(X)=0, D(X)=2. \quad 2. E(X)=\frac{81}{64}, \quad 3. \frac{5}{2}.$$

$$4. a = \frac{1}{4}, b = 1, c = -\frac{1}{4}, E(e^X) = \frac{1}{4}(e^2 - 1).$$

$$5. E(X)=16, D(X)=\frac{15}{4}. \quad 6. E(X)=\mu, D(X)=2\lambda^2.$$

第4章 随机向量

4.1 二维随机向量

4.1.1 知识要点

1. 定义

设 X, Y 是两个随机变量, 由它们构成的整体 ξ 称为二维随机向量, 记为 (X, Y) , 即 $\xi = (X, Y)$.

① 二维离散型随机向量: 如果二维随机向量 $\xi = (X, Y)$ 全部可能取到的值(二维向量)能够一一列举出来, 则称 ξ 为二维离散型随机向量.

设二维离散型随机向量 $\xi = (X, Y)$ 所有可能取到的值为

$$(x_i, y_j) \quad (i=1, 2, \dots; j=1, 2, \dots),$$

且取得各个值的概率为

$$p_{ij} = P[(X, Y) = (x_i, y_j)] \quad (i=1, 2, \dots; j=1, 2, \dots),$$

则称上式为 $\xi = (X, Y)$ 的联合概率分布, 简称为概率分布或联合分布.

二维离散型随机向量 $\xi = (X, Y)$ 的概率分布也可用概率分布表表示.

$P[(X, Y)]$		Y				
X		y_1	y_2	$\dots$	y_j	$\dots$
	x_1	p_{11}	p_{12}	$\dots$	p_{1j}	$\dots$
	x_2	p_{21}	p_{22}	$\dots$	p_{2j}	$\dots$
	$\vdots$	$\vdots$	$\vdots$			$\vdots$
	x_i	p_{i1}	p_{i2}	$\dots$	p_{ij}	$\dots$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	

p_{ij} 具有下列性质:

$$1^\circ p_{ij} \geq 0 \quad (i=1, 2, \dots; j=1, 2, \dots);$$

$$2^\circ \sum_i \sum_j p_{ij} = 1.$$

② 二维连续型随机向量: 对于二维随机向量 $\xi = (X, Y)$, 如果存在非负可积函数 $p(x, y)$

$(-\infty < x < +\infty, -\infty < y < +\infty)$ 使得对于任意一个由不等式 $a < x < b, c < y < d$ 确定的平面区域 $D = \{(x, y) | a < x < b, c < y < d\}$ 均有

$$P[(x, y) \in D] = \iint_D p(x, y) dx dy,$$

则称 $\xi = (X, Y)$ 为二维连续型随机向量, 且称 $p(x, y)$ 为 $\xi = (X, Y)$ 的联合分布密度, 简称为分布密度或(概率)密度.

$p(x, y)$ 具有下列性质:

$$1^\circ p(x, y) \geq 0;$$

$$2^\circ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = 1;$$

$$3^\circ \text{对于任意的平面区域 } D \text{ 均有 } P[(x, y) \in D] = \iint_D p(x, y) dx dy.$$

2. 二维随机向量的均匀分布

设 D 是平面上的有界区域, 其面积为 A . 如果二维随机向量 (X, Y) 的分布密度为

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

则称二维随机向量 (X, Y) 在区域 D 上服从均匀分布.

4.1.2 典型题型解析

例 4.1.1 一整数 X 随机地在 2, 3, 4 三个整数中取一个值, 另一整数 Y 随机地在 $2 \sim X$ 中取一值, 试求 (X, Y) 的联合分布.

解 X, Y 可能的取值为 2, 3, 4

$$P(X=2, Y=2) = \frac{1}{3} \times 1 = \frac{1}{3}, P(X=2, Y=k) = 0 \quad (k=3, 4),$$

$$P(X=3, Y=2) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}, P(X=3, Y=3) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6},$$

$$P(X=3, Y=4) = 0, P(X=4, Y=k) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9} \quad (k=2, 3, 4).$$

于是 (X, Y) 的联合概率分布为

P_{ij} \ Y				
		2	3	4
X	2	$\frac{1}{3}$	0	0
	3	$\frac{1}{6}$	$\frac{1}{6}$	0
	4	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

例 4.1.2 设 (X, Y) 的密度函数为

$$p(x, y) = \begin{cases} A \sin(x+y), & 0 < x < \frac{\pi}{2}, 0 < y < \frac{\pi}{2}, \\ 0, & \text{其他.} \end{cases}$$

求常数 A .

$$\begin{aligned} \text{解 } 1 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = A \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \sin(x+y) dx dy \\ &= A \int_0^{\frac{\pi}{2}} [-\cos(x+y)]_0^{\frac{\pi}{2}} dy = A \int_0^{\frac{\pi}{2}} (\sin y + \cos y) dy \\ &= A [-\cos y + \sin y]_0^{\frac{\pi}{2}} = 2A. \end{aligned}$$

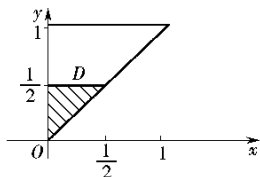
$$\text{所以 } A = \frac{1}{2}.$$

例 4.1.3 已知二维随机向量 (X, Y) 服从 $D = \{(x, y) | 0 < x < y < 1\}$ 上的均匀分布, 求 $P(0 < x < \frac{1}{2}, 0 < Y < \frac{1}{2})$.

解 区域 D 如右图所示.

所以区域 D 的面积为 $\frac{1}{2}$.

$$\text{因此 } p(x, y) = \begin{cases} 2, & (x, y) \in D, \\ 0, & \text{其他,} \end{cases}$$



$$\text{于是 } P\left(0 < X < \frac{1}{2}, 0 < Y < \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \int_x^{\frac{1}{2}} 2 dy dx = \frac{1}{4}.$$

同步练习 4.1

1. 在 10 件产品中有 2 件一级品, 7 件二级品和 1 件次品, 从 10 件产品中无放回地抽取 3 件, 用 X 表示其中的一级品数, Y 表示其中的二级品数, 求 (X, Y) 的联合概率分布.

2. 设随机向量 (X, Y) 的联合概率密度为

$$p(x, y) = \begin{cases} A e^{-(3x+4y)}, & x > 0, y > 0, \\ 0, & \text{其他,} \end{cases}$$

求: (1) 常数 A ; (2) $P(0 < x \leq 1, 0 < Y \leq 2)$.

3. 设 (X, Y) 在区域 D 上服从均匀分布, 其中 D 为直线 $y = x$ 和抛物线 $y = x^2$ 所围成的区域. 试求 (X, Y) 的联合分布密度.

同步练习 4.1 参考答案

1.		0	1	2	3
		0	0	$\frac{21}{120}$	$\frac{35}{120}$
	1	0	$\frac{14}{120}$	$\frac{42}{120}$	0
	2	$\frac{1}{120}$	$\frac{7}{120}$	0	0

$$2. A = 12 (e^{-3} - 1)(e^{-8} - 1). \quad 3. p(x, y) = \begin{cases} 6, & (x, y) \in D, \\ 0, & \text{其他.} \end{cases}$$

4.2 二维随机向量的分布

4.2.1 知识要点

1. 二维随机向量的分布函数

设二维随机向量为 (X, Y) . 记

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

则称二元函数 $F(x, y)$ 为二维随机向量 (X, Y) 的联合分布函数. 简称分布函数.

$F(x, y)$ 具有下列性质:

$$\textcircled{1} 0 \leq F(x, y) \leq 1 \quad (-\infty < x < +\infty, -\infty < y < +\infty);$$

$$\textcircled{2} \lim_{x \rightarrow -\infty} F(x, y) = 0, \lim_{y \rightarrow -\infty} F(x, y) = 0;$$

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

$$\textcircled{3} \text{当 } x_1 < x_2 \text{ 时, } F(x_1, y) \leq F(x_2, y);$$

$$\text{当 } y_1 < y_2 \text{ 时, } F(x, y_1) \leq F(x, y_2);$$

$$\textcircled{4} \text{对于二维离散型随机向量 } (X, Y), F(x, y) = \sum_{x_i \leq x, y_j \leq y} p_{ij};$$

$$\textcircled{5} \text{对于二维连续型随机向量 } (X, Y), F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(x, y) dx dy, \text{ 如果 } p(x, y) \text{ 连}$$

续, 则有 $\frac{\partial^2 F}{\partial x \partial y} = p(x, y);$

$$\textcircled{6} P(x_1 < X < x_2, y_1 < Y < y_2) = F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2).$$

2. 二维随机向量的边缘分布

$\textcircled{1}$ 设二维随机向量 (X, Y) 的联合分布函数为 $F(x, y)$. 如果记

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

则称 $F_X(x)$ 为 (X, Y) 关于 X 的边缘分布函数. 如果记

$$F_Y(y) = P(X < +\infty, Y \leq y) = \lim_{x \rightarrow +\infty} F(x, y),$$

则称 $F_Y(y)$ 为 (X, Y) 关于 Y 的边缘分布函数.

$$\textcircled{2} \text{对于离散型 } p_{i\cdot} = P(X = x_i) = \sum_{j=1}^{\infty} p_{ij} \quad (i = 1, 2, \dots),$$

$$p_{\cdot j} = P(Y = y_j) = \sum_{i=1}^{\infty} p_{ij} \quad (j = 1, 2, \dots),$$

分别称 $p_{i\cdot}$ 和 $p_{\cdot j}$ 为 (X, Y) 关于 X 和关于 Y 的边缘分布律.

$$\text{对于连续型 } p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy,$$

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x, y) dx,$$

分别称 $p_X(x)$ 和 $p_Y(y)$ 为 (X, Y) 关于 X 和关于 Y 的边缘概率密度.

3. 二维随机向量的独立性

设二维随机向量 (X, Y) 的联合分布函数及边缘分布函数分别为 $F(x, y)$ 及 $F_X(x)$ 、 $F_Y(y)$. 如果对于所有的 x, y 有

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

即 $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$,

则称随机变量 X 与 Y 是相互独立的.

对于离散型 X 与 Y 相互独立的充要条件为

$$P[(X, Y) = (x_i, y_j)] = P(X = x_i)P(Y = y_j),$$

即 $p_{ij} = p_{i\cdot} p_{\cdot j} \quad (i = 1, 2, \dots; j = 1, 2, \dots)$.

对于连续型 X 与 Y 相互独立的充要条件为 $p(x, y) = p_X(x)p_Y(y)$.

4. 二维正态分布

如果二维连续型随机向量 (X, Y) 的概率密度为

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

$(-\infty < x < +\infty, -\infty < y < +\infty; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho \text{ 均为常数, 且 } \sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1)$, 则称 (X, Y) 为服从参数 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 的二维正态分布, 且

$$p_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} \quad (-\infty < x < +\infty),$$

$$p_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2\sigma_2^2}(y-\mu_2)^2} \quad (-\infty < y < +\infty).$$

如果 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1, \sigma_2; \rho)$ 则 X 与 Y 相互独立 $\Leftrightarrow \rho = 0$.

4.2.2 典型题型解析

例 4.2.1 已知二维随机向量 (X, Y) 有联合分布函数

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

试求: (1) 边缘分布函数; (2) 联合分布密度; (3) X 与 Y 是否独立.

解 (1) $F_X(x) = \lim_{y \rightarrow +\infty} F(x, y) = 1 - e^{-x}, x > 0,$

$$F_Y(y) = \lim_{x \rightarrow +\infty} F(x, y) = 1 - e^{-y}, y > 0.$$

$$(2) p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = e^{-(x+y)}, x > 0, y > 0.$$

即 $p(x, y) = \begin{cases} e^{-(x+y)}, & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases}$

(3) 由于对于任意的 x, y 都有 $F(x, y) = F_X(x)F_Y(y)$ 成立, 故 X 与 Y 是相互独立的随

机变量.

注 这是一道常见的基本训练题. 对于二维连续型随机向量独立性的考察也可以从分布密度出发.

例 4.2.2 已知随机向量 (X, Y) 的概率分布如下:

P_{ij} \ Y		0	1	2	3
X					
0		0	k	$2k$	$3k$
1		$2k$	$3k$	$4k$	$5k$
2		$4k$	$5k$	$6k$	$7k$

试求: (1) $P\left(\frac{1}{2} \leq X \leq 2, Y < 3\right)$; (2) X 与 Y 是否独立?

解 (1) $1 = \sum_{i=1}^3 \sum_{j=1}^4 p_{ij} = 42k \Rightarrow k = \frac{1}{42}$,

$$\begin{aligned} P\left(\frac{1}{2} \leq X \leq 2, Y < 3\right) &= P(X=1, Y=0) + P(X=1, Y=1) + P(X=1, Y=2) + \\ &\quad P(X=2, Y=0) + P(X=2, Y=1) + P(X=2, Y=2) \\ &= 24k = \frac{24}{42} = \frac{4}{7}. \end{aligned}$$

(2) 为考察独立性, 先求边缘分布律.

X	0	1	2
$p_{i \cdot}$	$\frac{6}{42}$	$\frac{14}{42}$	$\frac{22}{42}$

,

Y	0	1	2	3
$p_{\cdot j}$	$\frac{6}{42}$	$\frac{9}{42}$	$\frac{12}{42}$	$\frac{15}{42}$

因为 $0 = p_{11} \neq p_{1 \cdot} \cdot p_{\cdot 1} = \frac{6}{42} \times \frac{6}{42}$, 所以 X 与 Y 不独立.

注 解答本题要利用离散型随机向量概率分布律的性质 1 解出 k , 才能往下做; 在求 $P\left(\frac{1}{2} \leq X \leq 2, Y < 3\right)$ 时, 要把所有情况考虑周全; 在考察独立性时只要有一个 $p_{ij} \neq p_{i \cdot} \cdot p_{\cdot j}$ 成立, 则不独立. 而只有当所有的 $p_{ij} = p_{i \cdot} \cdot p_{\cdot j}$ 成立时 X 与 Y 才独立 ($i=1, 2, 3$).

例 4.2.3 已知随机变量 X 和 Y 的概率分布分别为

X	-1	0	1
$p_{i \cdot}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

,

Y	0	1
$p_{\cdot j}$	$\frac{1}{2}$	$\frac{1}{2}$

且 $P(XY=0)=1$, 试求 X 和 Y 的联合概率分布.

解 因为 $P(XY=0)=1$, 所以有 $P(X=-1, Y=1)=P(X=1, Y=1)=0$,

即 $p_{12} = p_{32} = 0$.

则 X 和 Y 的联合分布律为

P_{ij} \ Y		0	1	$p_{i\cdot}$
X	-1	p_{11}	0	$\frac{1}{4}$
	0	p_{21}	p_{22}	$\frac{1}{2}$
	1	p_{31}	0	$\frac{1}{4}$
	$p_{\cdot j}$	$\frac{1}{2}$	$\frac{1}{2}$	

所以 $p_{11} = \frac{1}{4}$, $p_{31} = \frac{1}{4}$, $p_{22} = \frac{1}{2}$.

由于 $p_{21} + p_{22} = \frac{1}{2}$, 故 $p_{21} = \frac{1}{2} - \frac{1}{2} = 0$, 于是 (X, Y) 的联合分布律为

P_{ij} \ Y		0	1	$p_{i\cdot}$
X	-1	$\frac{1}{4}$	0	$\frac{1}{4}$
	0	0	$\frac{1}{2}$	$\frac{1}{2}$
	1	$\frac{1}{4}$	0	$\frac{1}{4}$
	$p_{\cdot j}$	$\frac{1}{2}$	$\frac{1}{2}$	1

注 利用对立事件可以确定 p_{12} , p_{32} , 这是关键. 随后充分利用联合分布律和边缘分布律的关系得到一系列等式, 从而确定 p_{ij} .

例 4.2.4 设二维随机变量 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$, 其概率密度为

$$p(x, y) = \frac{1}{2\pi\sqrt{3}} e^{-\frac{[4x^2 + 2xy + y^2 - 8x - 2y + 4]}{6}}, -\infty < x, y < +\infty$$

求参数 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 的值.

$$\text{解 因为 } \frac{4x^2 + 2xy + y^2 - 8x - 2y + 4}{6} = \frac{4(x-1)^2 + 2(x-1)y + y^2}{2(\sqrt{3})^2},$$

所以, 与二维正态分布的密度函数比较得

$$\mu_1 = 1, \mu_2 = 0, \sigma_1^2 = 1, \sigma_2^2 = 4, \rho = -\frac{1}{2}.$$

故知 $(X, Y) \sim N\left(1, 0; 1, 4; -\frac{1}{2}\right)$.

例 4.2.5 设二维随机向量 (X, Y) 的分布函数为

$$F(x, y) = A \left(B + \arctan \frac{x}{2} \right) \left(C + \arctan \frac{y}{3} \right), -\infty < x, y < +\infty.$$

求: (1) A, B, C 的值; (2) $p(x, y)$; (3) $p_X(x), p_Y(y)$.

解 (1) 因为 $A \neq 0$, 所以由 x, y 的任意性, 由

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

$$F(-\infty, 0) = A \left(B - \frac{\pi}{2} \right) \left(C + \arctan \frac{0}{3} \right) = 0, \text{ 得 } B = \frac{\pi}{2}.$$

$$F(+\infty, -\infty) = A \left(\frac{\pi}{2} + \frac{\pi}{2} \right) \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1, \text{ 得 } A = \frac{1}{\pi^2}.$$

所以 $F(x, y) = \frac{1}{\pi^2} \left(\frac{\pi}{2} + \arctan \frac{x}{2} \right) \left(\frac{\pi}{2} + \arctan \frac{y}{3} \right).$

(2) 由 $p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$ 得

$$p(x, y) = \frac{6}{\pi^2 (4+x^2)(9+y^2)}, -\infty < x, y < +\infty.$$

(3) $p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy = \frac{2}{\pi(4+x^2)}, -\infty < x < +\infty,$

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x, y) dx = \frac{3}{\pi(9+y^2)}, -\infty < y < +\infty.$$

注 利用分布函数的性质(2°)可以解出其中的未知常数,从而余下的问题迎刃而解.

例 4.2.6 设 X 与 Y 是相互独立的随机变量,且服从同一分布 $U[-a, a] (a > 0)$, 求方程 $t^2 + Xt + Y = 0$ 有实根的概率. 并求 $a \rightarrow 0$ 和 $a \rightarrow \infty$ 时, 此概率的极限值.

解 因为 $X \sim U[-a, a], Y \sim U[-a, a]$ 且 X 与 Y 相互独立, 所以

$$p(x, y) = \begin{cases} \frac{1}{4a^2}, & |x| \leq a, |y| \leq a, \\ 0, & \text{其他.} \end{cases}$$

要方程 $t^2 + Xt + Y = 0$ 有实根, 只要 $\Delta \geq 0$, 即 $X^2 - 4Y \geq 0$, 要求出

$$P(X^2 \geq 4Y) = \iint_{x^2 \geq 4y} p(x, y) dx dy.$$

当 $0 \leq a \leq 4$ 时, 由图(1)知, 有

$$\begin{aligned} P(X^2 \geq 4Y) &= 2 \int_0^a dx \int_{-\frac{x^2}{4}}^{\frac{x^2}{4}} \frac{1}{4a^2} dy \\ &= \frac{a}{24} + \frac{1}{2}, \end{aligned}$$

当 $a > 4$ 时, 由图(2)知, 有

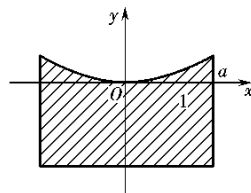
$$\begin{aligned} P(X^2 \geq 4Y) &= 1 - 2 \int_0^a dy \int_0^{\sqrt{4y}} \frac{1}{4a^2} dx \\ &= 1 - \frac{2}{3\sqrt{a}}, \end{aligned}$$

于是

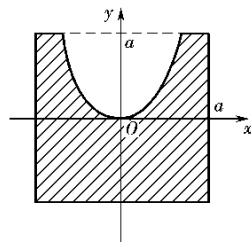
$$\lim_{a \rightarrow 0} P(X^2 \geq 4Y) = \lim_{a \rightarrow 0} \left(\frac{a}{24} + \frac{1}{2} \right) = \frac{1}{2},$$

$$\lim_{a \rightarrow \infty} P(X^2 \geq 4Y) = \lim_{a \rightarrow \infty} \left(1 - \frac{2}{3\sqrt{a}} \right) = 1.$$

注 首先利用独立性的定义得到 $p(x, y)$, 之后利用一元二次方程有实根的判别式 $\Delta \geq 0$ 得到所求概率为一二重积分. 注意分析 a 的不同取值范围得到的积分值是不一样的. 最后要正确地求出极限值.



图(1)



图(2)

同步练习 4.2

1. 设随机向量 (X, Y) 的分布函数为

$$F(x, y) = \begin{cases} c - 3^{-x} - 3^{-y} + 3^{-x-y}, & x \geq 0, y \geq 0, \\ 0, & \text{其他}, \end{cases}$$

求: (1) 常数 C ; (2) $p(x, y)$.

2. 设随机向量 (X, Y) 的联合分布律为

X \ Y	Y		
	0	1	2
0	0.06	0.15	α
1	β	0.35	0.21

问当 α, β 为何值时, X 与 Y 相互独立?

3. 设 X 与 Y 相互独立, $X \sim U[a, b], Y \sim \Gamma(1, \lambda)$ 求: (1) $p(x, y)$; (2) $P(Y \leq X)$.

4. 设二维随机向量 (X, Y) 的概率密度为

$$p(x, y) = A e^{-\frac{x^2 + y^2}{200}}, -\infty < x, y < +\infty.$$

求: (1) 常数 A ; (2) 确定 (X, Y) 的分布及分布的参数.

同步练习 4.2 参考答案

$$1. C = 1, p(x, y) = \begin{cases} 3^{-x-y} (\ln 3)^2, & x \geq 0, y \geq 0, \\ 0, & \text{其他}. \end{cases} \quad 2. \alpha = 0.09, \beta = 0.14.$$

$$3. p(x, y) = \begin{cases} \frac{\lambda e^{-\lambda y}}{b-a}, & a \leq x \leq b, y \geq 0, \\ 0, & \text{其他}. \end{cases} \quad P(Y \leq X) = 1 + \frac{1}{(b-a)\lambda} (e^{-\lambda b} - e^{-\lambda a}).$$

$$4. A = \frac{1}{200\pi} \quad (X, Y) \sim N(0, 0, 10^2, 10^2, 0).$$

4.3 二维随机向量函数的分布

4.3.1 知识要点

1. 定义

设 $f(x, y)$ 是定义在二维随机向量 (X, Y) 一切可能取值集合上的二元函数. 如果当二维随机向量 (X, Y) 取值 (x, y) 时, 随机变量 Z 就取值 $z = f(x, y)$, 则称 Z 为二维随机向量 (X, Y) 的函数, 记为 $Z = f(X, Y)$.

对于连续型 $F_Z(z) = P(Z \leq z) = P[f(X, Y) \leq z] = \iint_{f(x, y) \leq z} p(x, y) dx dy.$

2. 几种二维连续型随机向量函数

(1) $Z = X + Y$ 的分布

$$p_Z(z) = \int_{-\infty}^{+\infty} p(z-y, y) dy = \int_{-\infty}^{+\infty} p(x, z-x) dx.$$

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

(2) $Z = \max(X, Y)$ 的分布

设 X 与 Y 相互独立, 且 X 和 Y 的分布函数及密度分别为 $F_X(x)$, $F_Y(y)$ 及 $p_X(x)$, $p_Y(y)$.

$$F_{\max}(z) = P(Z \leq z) = F_X(z)F_Y(z).$$

$$p_{\max}(z) = p_X(z)F_Y(z) + F_X(z)p_Y(z).$$

(3) $Z = \sqrt{X^2 + Y^2}$ 的分布

$$F_Z(z) = \iint_{x^2 + y^2 \leq z^2} f(x, y) dx dy.$$

4.3.2 典型题型解析

例 4.3.1 设二维随机向量 (X, Y) 有联合概率密度

$$p(x, y) = \begin{cases} 2xe^{-y}, & 0 < x < 1, y > 0, \\ 0, & \text{其他}, \end{cases}$$

试求 $Z = X + Y$ 的概率密度.

解 此处随机变量 Z 是任意随机变量的和, 故使用公式

$$p_Z(z) = \int_{-\infty}^{+\infty} p(x, z-x) dx$$

来求 Z 的概率密度. 由于参变量 z 可以取一切实数, 故要在 z 的不同取值范围内进行逐段积分. 为方便, 下面的讨论借助右图进行:

$$(1) \text{ 当 } z < 0 \text{ 时, 有 } p_Z(z) = \int_{-\infty}^{+\infty} p(x, z-x) dx = 0;$$

$$(2) \text{ 当 } 0 \leq z < 1 \text{ 时, 有 } p_Z(z) = \int_{-\infty}^{+\infty} p(x, z-x) dx,$$

由于 $y = z - x > 0 \Rightarrow x < z$
 又由题设知 $0 < x < 1$ } $\Rightarrow 0 < x < z$,

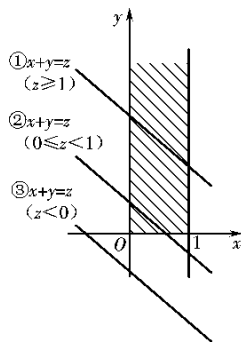
$$\text{故 } p_Z(z) = \int_0^z 2xe^{-(z-x)} dx = 2(e^{-z} + z - 1);$$

$$(3) \text{ 当 } z \geq 1 \text{ 时, 有 } p_Z(z) = \int_{-\infty}^{+\infty} p(x, z-x) dx,$$

由于 $y = z - x > 0 \Rightarrow x < z$
 又由题设知 $0 < x < 1$ } $\Rightarrow 0 < x < 1$,

$$\text{故 } p_Z(z) = \int_0^1 2xe^{-(z-x)} dx = 2e^{-z}.$$

综上, 可得 $Z = X + Y$ 的概率密度为



$$p_Z(z) = \begin{cases} 0, & z < 0, \\ 2(e^{-z} + z - 1), & 0 \leq z < 1, \\ 2e^{-z}, & z \geq 1. \end{cases}$$

注 本题是二维随机向量和分布的典型例子. 对于参变量 z 的不同取值范围, 其积分上下限是有差别的. 解题中提供的处理方法有普遍意义, 务求理解, 以便举一反三地求解这类题目. 此外, 应注意到题中随机变量 X 和 Y 实际上是相互独立的. 因而也可以使用卷积公式, 求解结果一致.

例 4.3.2 设随机变量 X, Y 相互独立, 都在区间 $[a, b]$ 上服从均匀分布, 求: (1) $Z_1 = \max(X, Y)$ 的概率密度; (2) $Z_2 = \min(X, Y)$ 的概率密度.

解 这里 X 与 Y 相互独立, 都服从 $[a, b]$ 上的均匀分布.

(1) Z_1 的分布函数为

$$\begin{aligned} F_{Z_1}(z) &= P(Z_1 \leq z) = P[\max(X, Y) \leq z] P(X \leq z, Y \leq z) \\ &= P(X \leq z) P(Y \leq z). \end{aligned}$$

$$\text{当 } a \leq z \leq b \text{ 时, } F_{Z_1}(z) = \frac{z-a}{b-a} \cdot \frac{z-a}{b-a} = \left(\frac{z-a}{b-a}\right)^2,$$

$$\text{当 } z < a \text{ 时, } F_{Z_1}(z) = 0,$$

$$\text{当 } z > b \text{ 时, } F_{Z_1}(z) = 1.$$

所以

$$F_{Z_1}(z) = \begin{cases} 0, & z < a, \\ \left(\frac{z-a}{b-a}\right)^2, & a \leq z \leq b, \\ 1, & z > b. \end{cases}$$

从而 Z_1 的概率密度为

$$p_{Z_1}(z) = \begin{cases} \frac{2(z-a)}{(b-a)^2}, & a \leq z \leq b, \\ 0, & \text{其他.} \end{cases}$$

(2) Z_2 的分布函数为

$$\begin{aligned} F_{Z_2}(z) &= P(Z_2 \leq z) = 1 - P(Z_2 > z) = 1 - P[\min(X, Y) > z] \\ &= 1 - P(X > z, Y > z) = 1 - P(X > z) P(Y > z) = 1 - [1 - P(X \leq z)]^2 \\ &= \begin{cases} 0, & z < a, \\ 1 - \left(1 - \frac{z-a}{b-a}\right)^2, & a \leq z \leq b, \\ 1, & z > b. \end{cases} \end{aligned}$$

从而 Z_2 的概率密度为

$$p_{Z_2}(z) = \begin{cases} \frac{2(b-z)}{(b-a)^2}, & a \leq z \leq b, \\ 0, & \text{其他.} \end{cases}$$

注 在使用 $Z = \max(X, Y)$ 的有关公式时, 要注意这些公式对 X, Y 是有要求的, 仅当 X

与 Y 相互独立时才适用.

例 4.3.3 设二维随机向量 (X, Y) 在矩形域 $G = \{(x, y) | 0 \leq x \leq 2, 0 \leq y \leq 1\}$ 上服从均匀分布, 试求边长为 X 和 Y 的矩形面积 S 的概率密度.

解 二维随机向量 (X, Y) 的联合概率密度为

$$p(x, y) = \begin{cases} \frac{1}{2}, & (x, y) \in G, \\ 0, & (x, y) \notin G. \end{cases}$$

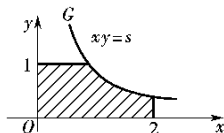
设 $S = XY$, 则 S 的分布函数为

$$F(s) = P(S \leq s) = P(XY \leq s) = \iint_{xy \leq s} p(x, y) dx dy = \iint_G p(x, y) dx dy.$$

显然, 当 $s < 0$ 时, $F(s) = 0$; 当 $s \geq 2$ 时, $F(s) = 1$.

当 $0 \leq s < 2$ 时, 积分区域 G 如右图所示, 从而

$$\begin{aligned} F(s) &= \int_0^s dx \int_0^1 \frac{1}{2} dy + \int_s^2 dx \int_0^{\frac{s}{x}} \frac{1}{2} dy \\ &= \int_0^s \frac{1}{2} dx + \int_s^2 \frac{s}{2x} dx = \frac{s}{2} (1 + \ln 2 - \ln s). \end{aligned}$$



所以

$$F(s) = \begin{cases} 0, & s < 0, \\ \frac{s}{2} (1 + \ln 2 - \ln s), & 0 \leq s < 2, \\ 1, & s \geq 2, \end{cases}$$

于是, S 的密度函数为

$$p(s) = \begin{cases} \frac{1}{2} (\ln 2 - \ln s), & 0 < s < 2, \\ 0, & \text{其他.} \end{cases}$$

同步练习 4.3

1. 设随机变量 X, Y 独立, $X \sim N(\mu, \sigma^2)$, Y 服从 $[-\pi, \pi]$ 上的均匀分布, 试求 $Z = X + Y$ 的概率密度函数.

2. 设某系统由 3 个相互独立的元件 A_1, A_2, A_3 连接而成, 连接方式为: (1) 串联, (2) 并联. 设 A_1, A_2, A_3 的寿命分别为随机变量 X_1, X_2, X_3 , 它们有相同的分布, 都服从指数分布:

$$p_X(x) = \begin{cases} a e^{-ax}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

试分别按上述两种连接方式, 求出系统 (1), (2) 的寿命的概率密度和平均寿命.

3. 设二维随机向量 (X, Y) 的概率密度为

$$p(x, y) = \begin{cases} 2e^{-(x+2y)}, & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases}$$

求随机变量 $Z = X + 2Y$ 的分布函数.

同步练习 4.3 参考答案

$$1. p_z(z) = \frac{1}{2\pi} \left[\Phi\left(\frac{z + \pi - \mu}{\sigma}\right) - \Phi\left(\frac{z - \pi - \mu}{\sigma}\right) \right];$$

$$2. (1) p_{Z_1}(z) = \begin{cases} 3ae^{-3az}, & z > 0, \\ 0, & z \leq 0, \end{cases} \quad E(Z_1) = \frac{1}{3a};$$

$$(2) p_{Z_2}(z) = \begin{cases} 3(1 - e^{-az})^2 \cdot ae^{-az}, & z > 0, \\ 0, & z \leq 0, \end{cases} \quad E(Z_2) = \frac{33}{18a};$$

$$3. F_Z(z) = \begin{cases} 1 - e^{-z} - ze^{-z}, & z > 0, \\ 0, & z \leq 0. \end{cases}$$

4.4 二维随机向量的数字特征

4.4.1 知识要点

1. 二维随机向量的数字特征

设二维随机向量 (X, Y) 的联合密度为 $p(x, y)$, $Z = f(X, Y)$, 关于 X, Y 的边缘密度分别为 $p_X(x), p_Y(y)$. 则有

$$(1) E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) p(x, y) dx dy;$$

$$(2) E(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x p(x, y) dx dy;$$

$$(3) E(Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y p(x, y) dx dy;$$

$$(4) D(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - E(X)]^2 p(x, y) dx dy;$$

$$(5) D(Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [y - E(Y)]^2 p(x, y) dx dy;$$

$$(6) \text{cov}(X, Y) = \sigma_{XY} = E\{[X - E(X)][Y - E(Y)]\}.$$

特别是对于二维连续型随机向量

$$\sigma_{XY} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - E(X)][y - E(Y)] p(x, y) dx dy,$$

且 $\sigma_{XX} = D(X), \sigma_{YY} = D(Y);$

$$(7) \rho = \rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}.$$

2. 数字特征的简单性质

$$(1) E(X \pm Y) = E(X) \pm E(Y);$$

$$(2) \text{如果 } X \text{ 与 } Y \text{ 相互独立, 则有 } E(XY) = E(X)E(Y);$$

$$(3) D(X \pm Y) = D(X) + D(Y) \pm 2\sigma_{XY};$$

$$(4) \text{如果 } X \text{ 与 } Y \text{ 相互独立, 则有 } D(X \pm Y) = D(X) + D(Y);$$

$$(5) \text{如果 } X \text{ 与 } Y \text{ 相互独立, 则有 } \sigma_{XY} = 0 (\text{或 } \rho_{XY} = 0);$$

$$(6) |\rho_{XY}| \leq 1.$$

4.4.2 典型题型解析

例 4.4.1 已知随机向量 (X, Y) 联合密度为

$$p(x, y) = \begin{cases} 4xy, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{其他.} \end{cases}$$

试求 $E(X^2 + Y^2)$.

$$\begin{aligned} \text{解 } E(X^2 + Y^2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^2 + y^2) p(x, y) dx dy \\ &= \int_0^1 \int_0^1 (x^2 + y^2) 4xy dx dy \\ &= 4 \int_0^1 x^3 dx \int_0^1 y dy + 4 \int_0^1 x dx \int_0^1 y^3 dy \\ &= \frac{1}{2} + \frac{1}{2} = 1. \end{aligned}$$

例 4.4.2 设二维随机向量 (X, Y) 联合密度为

$$p(x, y) = \begin{cases} \frac{3xy}{16}, & (x, y) \in G, \\ 0, & (x, y) \notin G, \end{cases}$$

其中 G 为 $0 \leq x \leq 2$ 及 $0 \leq y \leq x^2$ 所围的区域. 试求 $E(X), E(Y), D(X), D(Y), \text{cov}(X, Y)$

ρ_{XY} .

解 先求边缘概率密度

$$\begin{aligned} p_X(x) &= \int_0^{x^2} \frac{3xy}{16} dy = \frac{3x^5}{32}, \quad 0 \leq x \leq 2, \\ p_Y(y) &= \int_{\sqrt{y}}^2 \frac{3xy}{16} dx = \frac{3y(4-y)}{32}, \quad 0 \leq y \leq 4, \end{aligned}$$

于是 $E(X) = \int_0^2 x \cdot \frac{3}{32} x^5 dx = \frac{3}{32} \int_0^2 x^6 dx = \frac{12}{7},$

$$E(X^2) = \int_0^2 x^2 \cdot \frac{3}{32} x^5 dx = \frac{3}{32} \int_0^2 x^7 dx = 3,$$

$$D(X) = E(X^2) - (EX)^2 = 3 - \left(\frac{12}{7}\right)^2 = \frac{3}{49},$$

$$E(Y) = \int_0^4 y \cdot \frac{3}{32} y(4-y) dy = \frac{3}{32} \int_0^4 (4y^2 - y^3) dy = 2,$$

$$E(Y^2) = \int_0^4 y^2 \cdot \frac{3}{32} y(4-y) dy = \frac{3}{32} \int_0^4 (4y^3 - y^4) dy = \frac{24}{5},$$

$$D(Y) = E(Y^2) - (EY)^2 = \frac{24}{5} - 4 = \frac{4}{5}.$$

又因为 $E(XY) = \int_0^2 dx \int_0^{x^2} xy \frac{3}{16} xy dy = \frac{1}{16} \int_0^2 x^8 dx = \frac{32}{9}$,

故 $\text{cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = \frac{32}{9} - \frac{24}{7} = \frac{8}{63}$,

于是 $\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{\frac{8}{63}}{\sqrt{\frac{3}{49}}\sqrt{\frac{4}{5}}} = \frac{4}{9}\sqrt{\frac{5}{3}}$.

例 4.4.3 已知随机变量 X, Y, Z , 证明:

$D(X+Y+Z) = D(X) + D(Y) + D(Z) + 2\text{cov}(X, Y) + 2\text{cov}(X, Z) + 2\text{cov}(Y, Z)$:

证明 $D(X+Y+Z) = E\{[(X+Y+Z) - E(X+Y+Z)]^2\}$
 $= E\{[X - E(X)] + [Y - E(Y)] + [Z - E(Z)]\}^2\}$
 $= E\{[X - E(X)]^2\} + E\{[Y - E(Y)]^2\} + E\{[Z - E(Z)]^2\} +$
 $2E\{[X - E(X)] \cdot [Y - E(Y)]\} + 2E\{[X - E(X)] \cdot [Z - E(Z)]\} +$
 $2E\{[Y - E(Y)] \cdot [Z - E(Z)]\}$
 $= D(X) + D(Y) + D(Z) + 2\text{cov}(X, Y) + 2\text{cov}(X, Z) + 2\text{cov}(Y, Z).$

例 4.4.4 已知 $X \sim N(1, 3^2)$, $Y \sim N(0, 4^2)$, $\rho_{XY} = -\frac{1}{2}$, 对 $Z = \frac{X}{3} + \frac{Y}{2}$, 求: (1) $E(Z)$ 和 $D(Z)$; (2) ρ_{XZ} ; (3) X, Z 的独立性.

解 (1) 由分布知 $E(X) = 1, D(X) = 9, E(Y) = 0, D(Y) = 16$, 所以

$$E(Z) = \frac{1}{3}E(X) + \frac{1}{2}E(Y) = \frac{1}{3} + 0 = \frac{1}{3},$$

$$D(Z) = \frac{1}{9}D(X) + \frac{1}{4}D(Y) + 2 \cdot \frac{1}{3} \cdot \frac{1}{2} \text{cov}(X, Y)$$

$$= 1 + 4 + \frac{1}{3} \left(-\frac{1}{2} \right) \times 4 \times 3 = 3.$$

$$(2) \text{cov}(X, Z) = \text{cov}\left(X, \frac{X}{3} + \frac{Y}{2}\right)$$

$$= \frac{1}{3}\text{cov}(X, X) + \frac{1}{2}\text{cov}(X, Y)$$

$$= \frac{1}{3}D(X) + \frac{1}{2} \left(-\frac{1}{2} \right) \times 3 \times 4 = 3 - 3 = 0,$$

所以 $\rho_{XZ} = 0$.

(3) 因为 $Z = \frac{X}{3} + \frac{Y}{2}$, 所以 Z 也是正态分布.

故由 $\rho_{XZ} = 0$ 知 X, Z 相互独立.

注 对于二维正态分布来说, $\rho_{XZ} = 0$ 是 X 与 Z 相互独立的充分必要条件.

例 4.4.5 设随机变量 X 与 Y 独立同服从 $N(\mu, \sigma^2)$ 分布. 令

$$Z = |X - Y|, U = \max(X, Y), V = \min(X, Y).$$

求: (1) $E(Z), D(Z)$; (2) $E(U), E(V)$.

解 (1) 由 $X \sim N(\mu, \sigma^2), Y \sim N(\mu, \sigma^2)$ 得

$$E(X)=E(Y)=\mu, D(X)=D(Y)=\sigma^2,$$

于是 $E(X-Y)=E(X)-E(Y)=0$.

由 X 与 Y 的独立性得

$$D(X-Y)=D(X)+D(Y)=2\sigma^2,$$

所以 $\frac{X-Y}{\sqrt{2\sigma^2}} \sim N(0, 1)$.

$$\text{令 } T = \frac{X-Y}{\sqrt{2}\sigma} \sim N(0, 1),$$

则 $Z = (\sqrt{2}\sigma)|T|$,

$$\begin{aligned} \text{故 } E(Z) &= E[(\sqrt{2}\sigma)|T|] = \sqrt{2}\sigma \int_{-\infty}^{+\infty} |t| \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= \frac{2\sigma}{\sqrt{\pi}} \int_0^{+\infty} t e^{-\frac{t^2}{2}} dt = \frac{2\sigma}{\sqrt{\pi}} \left(-e^{-\frac{t^2}{2}} \right) \Big|_0^{+\infty} = \frac{2\sigma}{\sqrt{\pi}}. \end{aligned}$$

又因为 $E(T^2)=D(T)+(ET)^2=1$,

所以 $E(Z^2)=E[(2\sigma^2)T^2]=2\sigma^2 E(T^2)=2\sigma^2$.

$$D(Z)=E(Z^2)-(EZ)^2=2\sigma^2-\left(\frac{2\sigma}{\sqrt{\pi}}\right)^2=2\sigma^2\left(1-\frac{2}{\pi}\right).$$

$$(2) \text{ 因为 } \max(X, Y) = \frac{1}{2}(X+Y+|X-Y|),$$

$$\min(X, Y) = \frac{1}{2}(X+Y-|X-Y|),$$

所以

$$\begin{aligned} E[\max(X, Y)] &= \frac{1}{2}[EX+EY+E(|X-Y|)] \\ &= \frac{1}{2}[2\mu+E(Z)] = \frac{1}{2}\left(2\mu+\frac{2\sigma}{\sqrt{\pi}}\right) = \mu + \frac{\sigma}{\sqrt{\pi}}. \end{aligned}$$

$$\begin{aligned} E[\min(X, Y)] &= \frac{1}{2}[EX+EY-E(|X-Y|)] \\ &= \frac{1}{2}[2\mu-E(Z)] = \mu - \frac{\sigma}{\sqrt{\pi}}. \end{aligned}$$

同步练习 4.4

1. 设随机向量 (X, Y) 的概率密度为

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

求: (1) $f_X(x)f_Y(y)$ (2) $E(X), E(Y), D(X), D(Y)$ (3) $\text{cov}(X, Y)$.

2. 设 X 和 Y 为两个随机变量, $E(X)=E(Y)=0, D(X)=D(Y)=1, \rho_{XY}=\text{cov}(X, Y)$.

证明: $E[\max(X^2, Y^2)] \leq 1 + \sqrt{1 - \rho^2}$

3. 设随机向量 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$, 其中 $\mu_1 = 1, \mu_2 = -2; \sigma_1^2 = 16, \sigma_2^2 = 9$,

$\rho = \frac{1}{2}$. 试求: (1) $E(X - Y + 1)$; (2) $D(X - Y + 1)$; (3) $E(X - Y)^2$.

同步练习 4.4 参考答案

$$1. f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0, & \text{其他}, \end{cases} \quad f_Y(y) = \begin{cases} 1 - |y|, & -1 \leq y \leq 1, \\ 0, & \text{其他}, \end{cases}$$

$$E(X) = \frac{2}{3}, D(X) = \frac{1}{18}, E(Y) = 0, D(Y) = \frac{1}{6}, \text{cov}(X, Y) = 0.$$

2. 利用数学期望与方差的关系式, 特别利用了 $E(X) = 0, E(Y) = 0, E(X + Y) = 0$.

3. $E(X - Y + 1) = 4, D(X - Y + 1) = 13, E(X - Y)^2 = 22$.

4.5 大数定律与中心极限定理

4.5.1 知识要点

1. n 维随机向量简介

① 设 $X_1, X_2, \dots, X_n$ 是 n 个随机变量, 由它们构成的整体 ξ 称为 n 维随机向量, 记为 $(X_1, X_2, \dots, X_n)$. 即 $\xi = (X_1, X_2, \dots, X_n)$;

② 对于 n 维随机向量 $\xi = (X_1, X_2, \dots, X_n)$, 如果存在非负可积函数 $p(x_1, x_2, \dots, x_n)$ ($-\infty < x_i < +\infty, i = 1, 2, \dots, n$) 使得对于任意一个由不等式 $a_i < x_i < b_i (i = 1, 2, \dots, n)$ 确定的区域 $D = \{(x_1, \dots, x_n) | a_1 < x_1 < b_1, \dots, a_n < x_n < b_n\}$ 均有 $P[(X_1, X_2, \dots, X_n) \in D] = \iint_D \dots \int p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$, 则称 $\xi = (X_1, X_2, \dots, X_n)$ 为 n 维连续型随机向量, 而称 $p(x_1, x_2, \dots, x_n)$ 为 $\xi = (X_1, X_2, \dots, X_n)$ 的联合分布密度;

③ n 维连续型随机向量 $(X_1, X_2, \dots, X_n)$ 的一部分分量所构成的随机向量 $(X_{i_1}, X_{i_2}, \dots, X_{i_k}) (1 \leq i_1 < i_2 < \dots < i_k \leq n)$ 的分布密度称为 $(X_1, X_2, \dots, X_n)$ 的边缘密度. 特别地, 每个分量 X_i 的分布密度 $p_i(x_i) (i = 1, 2, \dots, n)$ 也称为 $(X_1, X_2, \dots, X_n)$ 的边缘密度, 且称为 $(X_1, X_2, \dots, X_n)$ 的单个密度;

④ 如果恒有 $p(x_1, x_2, \dots, x_n) = p_1(x_1)p_2(x_2)\dots p_n(x_n)$, 则称 $X_1, X_2, \dots, X_n$ 是相互独立的;

⑤ 如果记 $F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$, 则称 n 元函数 $F(x_1, x_2, \dots, x_n)$ 为 n 维随机向量 $(X_1, X_2, \dots, X_n)$ 的联合分布函数;

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n.$$

如果 $p(x_1, x_2, \dots, x_n)$ 连续, 则有

$$\frac{\partial^n F}{\partial x_1 \partial x_2 \dots \partial x_n} = p(x_1, x_2, \dots, x_n).$$

⑥ 设随机变量列为 $X_1, X_2, \dots, X_n, \dots$. 如果对于任何 $n > 1$, $X_1, X_2, \dots, X_n$ 都是相互独立的, 且所有的 $X_i (i = 1, 2, \dots)$ 又有共同的分布函数, 则称 $X_1, X_2, \dots, X_n, \dots$ 是独立同分布的随机变量列;

⑦ n 维连续型随机向量 $(X_1, X_2, \dots, X_n)$ 的函数 $Y = f(X_1, X_2, \dots, X_n)$ 的均值公式为

$$E[f(X_1, X_2, \dots, X_n)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

2. 性质

① $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$;

② 当 $X_1, X_2, \dots, X_n$ 相互独立时, 有

$$E(X_1 X_2 \dots X_n) = E(X_1) E(X_2) \dots E(X_n);$$

③ $D(X_1 + X_2 + \dots + X_n) = D(X_1) + D(X_2) + \dots + D(X_n)$.

3. 大数定律

设 $X_1, X_2, \dots, X_n, \dots$ 是独立同分布的随机变量列, 且 $E(X_1), D(X_1)$ 存在, 则对任何正数 ϵ 均有

$$\lim_{n \rightarrow \infty} P\left[\left|\frac{S_n}{n} - E(X_1)\right| \geq \epsilon\right] = 0 \quad (S_n = X_1 + X_2 + \dots + X_n).$$

4. 中心极限定理

设 $X_1, X_2, \dots, X_n, \dots$ 是独立同分布的随机变量列, 且 $E(X_1), D(X_1)$ 存在, $D(X_1) \neq 0$, 则对任意的实数 $a < b$ 均有

$$\lim_{n \rightarrow \infty} P\left[a < \frac{S_n - nE(X_1)}{\sqrt{nD(X_1)}} < b\right] = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (S_n = X_1 + X_2 + \dots + X_n).$$

4.5.2 典型题型解析

例 4.5.1 设 $X_1, X_2, \dots, X_{n+m} (n > m)$ 是独立同分布且方差存在的随机变量, 又令 $Y = X_1 + X_2 + \dots + X_n, Z = X_{m+1} + X_{m+2} + \dots + X_{m+n}$, 求 ρ_{YZ} .

解 为简单计, 不妨设 $E(X_i) = 0, E(X_i^2) = 1, i = 1, 2, \dots, m+n$, 则

$$\begin{aligned} \text{cov}(Y, Z) &= E[(X_1 + X_2 + \dots + X_n)(X_{m+1} + X_{m+2} + \dots + X_{m+n})] \\ &= E[X_{m+1}^2 + X_{m+2}^2 + \dots + X_n^2] = n - m, \end{aligned}$$

$$D(Y) = E(X_1^2 + X_2^2 + \dots + X_n^2) = n,$$

$$D(Z) = E(X_{m+1}^2 + X_{m+2}^2 + \dots + X_{m+n}^2) = n,$$

故
$$\rho_{YZ} = \frac{\text{cov}(Y, Z)}{\sqrt{D(Y)}\sqrt{D(Z)}} = \frac{n-m}{n} = 1 - \frac{m}{n}.$$

注 为方便计算, 设 $E(X_i) = 0, E(X_i^2) = 1 (i = 1, 2, \dots, m+n)$, 若设成其他的值结果也一样, 只是计算起来比较麻烦.

例 4.5.2 已知独立随机变量序列 $X_1, X_2, \dots, X_n$ 具有同一分布

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{a}\right).$$

问:是否可以适用大数定律?

解 因为 $p(x) = \frac{a}{\pi(a^2 + x^2)}$, 所以

$$E(X) = \int_{-\infty}^{+\infty} |x| \frac{a}{\pi(a^2 + x^2)} dx = \frac{2a}{\pi} \int_0^{+\infty} \frac{ax}{a^2 + x^2} dx = \infty,$$

所以 $E(X)$ 不存在. 而大数定律要求 $E(X)$ 存在, 故该随机变量序列不适合大数定律.

注 验证是否满足大数定律的条件时, 要判断 $E(X)$ 和 $D(X)$ 的存在性, 同时要用到计算数学期望和方差的技巧.

例 4.5.3 设一条自动生产线的产品合格率是 0.8, 要使一批产品的合格率在 76% 与 84% 之间的概率不小于 90%. 问这批产品至少要生产多少件?

解 用中心极限定理估计, 可知当 m 比较大时, X 近似服从正态分布 $N(0.8m, 0.16m)$, 于是, 有

$$P\left\{0.76 < \frac{X}{m} < 0.84\right\} = P\left\{\left|\frac{X - 0.8m}{0.4\sqrt{m}}\right| < \frac{0.04m}{0.4\sqrt{m}}\right\} \approx 2\Phi(0.1\sqrt{m}) - 1 \geq 0.90,$$

查正态分布表可知 $0.1\sqrt{m} = 1.65$, 解得 $m \geq 268.96$, 故取 $m = 269$.

注 利用中心极限定理讨论的一类问题是在概率确定的条件下求样本数 m . 解题方法是确定随机变量序列, 建立标准化随机变量, 建立已知概率与标准正态分布的关系式, 寻找 m 的表达式, 通过正态分布表得出 m 的关系式即可求出 m .

同步练习 4.5

1. 设 $X_1, X_2, \dots, X_n$ 独立同分布, 都服从 $N(\mu, \sigma^2)$, 试求 $E(X_1 X_2 \dots X_n)$.
2. 设有如下随机变量序列, 两两独立

X_k	$-\sqrt{2}$	0	$\sqrt{2}$
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

问: 可否对此随机变量序列使用大数定律?

3. 某单位有 1 000 人各自独立地参加防空演习, 设每个人能按时进入掩体的概率为 0.9. 以 0.95 的概率估计: 在一次演习中至少有多少人能进入掩体?

同步练习 4.5 参考答案

1. μ^n .
2. 可以使用.
3. 884 人.

疑难解析

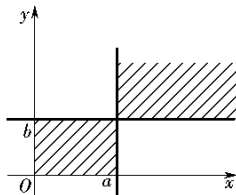
1. 事件 $P\{X \leq x, Y \leq y\}$ 表示事件 $\{X \leq x\}$ 与 $\{Y \leq y\}$ 的积事件, 为什么 $P\{X \leq x, Y \leq y\}$ 不一定等于 $P\{X \leq x\}P\{Y \leq y\}$?

答 与仅当事件 A, B 相独立时, 才有 $P(AB) = P(A)P(B)$ 一样. 这里, 依乘法原理有 $P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y | X \leq x\}$, 那么, 当 $P\{X \leq x\}$ 和 $P\{Y \leq y\}$ 相互独立时, 有 $P\{Y \leq y | X \leq x\} = P\{Y \leq y\}$, 因而下式成立:

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}.$$

2. 事件 $\{X \leq a, Y \leq b\}$ 与 $\{X > a, Y > b\}$ 是否为对立事件,为什么?

答 事件 $\{X \leq a, Y \leq b\}$ 与 $\{X > a, Y > b\}$ 不是对立事件.由右图知,事件 $\{X \leq a, Y \leq b\}$ 发生,即随机点 (X, Y) 落在图中左下部阴影区域内;事件 $\{X > a, Y > b\}$ 发生,即随机点 (X, Y) 落在图中右上部阴影区域内.它们的和事件不遮盖全面区域,所以不是对立事件.



3. 计算概率 $P\{(X, Y) \in G\}$ 时应注意些什么?

答 当 (X, Y) 是离散型随机变量时, $P\{(X, Y) \in G\} = \sum_{(x_i, y_j) \in G} p_{ij}$,注意必须找出 G 内所有使 $p_{ij} \neq 0$ 的点 (x_i, y_j) ,不能遗漏.

当 (X, Y) 是连续型随机变量时, $P\{(X, Y) \in G\} = \iint_G f(x, y) dx dy$ 注意必须正确确定二重积分的积分限.特别是在区域 G 要分块计算时,不要有遗漏或重复求积发生.

在求 (X, Y) 的分布函数 $F(x, y)$ 时,对 $\mathbf{R}^2$ 平面上的各个区域的概率都要正确求出,并按顺序累积.

4. 若 X, Y 相互独立,则 $f_1(X)$ 与 $f_2(Y)$ 相互独立, f_1, f_2 均为连续函数.试问逆命题是否正确?

答 逆命题不正确.例如:设 (X, Y) 的联合密度函数为

$$p(x, y) = \begin{cases} \frac{1+xy}{4}, & |x| < 1, |y| < 1, \\ 0, & \text{其他,} \end{cases}$$

则 X 与 Y 不独立,但 X^2 与 Y^2 独立(留给读者自己证明).但若 f_1, f_2 存在反函数时逆命题成立.

5. 边缘分布均为正态分布的随机变量,其联合分布一定是二维正态分布吗?

答 不一定.例如:设 (X, Y) 的联合密度函数为

$$p(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} (1 + \sin x \sin y),$$

显然 (X, Y) 不服从正态分布,但是有 $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$.

此例说明,边缘分布都为正态分布,但其联合分布却不是正态分布.

由此可知,一般地联合密度能决定边缘密度,但边缘密度并不能决定联合密度,只有在特殊情况下,即当 X, Y 相互独立时,才有 $f(x, y) = f_X(x)f_Y(y)$.

6. 设 X, Y 相互独立,且都服从参数为 p 的两点分布,则一定有 $X = Y$ 吗?

答 不对,因为这是误认为 X, Y 只能取0,1两个值,而取0,1的概率都是 $1-p$ 和 p ,因此 $X = Y$ 是错误的.事实上, X, Y 都是“随机”取值的,当 X 取0时, Y 可能取0也可能取1,而同时取0或同时取1的可能值为 $(1-p)^2 + p^2 = 1 - 2p + 2p^2 < 1$.

7. 若 (X, Y) 的联合分布函数 $F(x, y)$ 在 (x_0, y_0) 连续,则 X 和 Y 的边缘分布函数 $F_X(x)$ 和 $F_Y(y)$ 在点 x_0 和 y_0 一定连续吗?

答 不一定,例如,设 (X, Y) 的分布律为

		Y	
		0	1
X	0	0	$\frac{1}{2}$
	1	$\frac{1}{2}$	0

则 (X, Y) 的联合分布函数

$$F(x, y) = \begin{cases} 0, & x < 1 \text{ 且 } y < 1, \\ \frac{1}{2}, & x \geq 1 \text{ 且 } y < 1, \\ \frac{1}{2}, & x < 1 \text{ 且 } y \geq 1, \\ 1, & x \geq 1 \text{ 且 } y \geq 1, \end{cases}$$

显然 $F(x, y)$ 在 $(0, 0)$ 连续, 但 X 的边缘分布函数

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

在 $x=0$ 间断.

$$\text{同理, } Y \text{ 的边缘分布函数 } F_Y(y) = \begin{cases} 0, & y < 0, \\ \frac{1}{2}, & 0 \leq y < 1, \\ 1, & y \geq 1, \end{cases}$$

在 $y=0$ 间断.

8. 边缘分布与联合分布的关系如何?

答 (X, Y) 的联合分布全面反映了 (X, Y) 的概率分布状态以及数字特征等, 而边缘分布只反映分量 X 或 Y 的概率分布, 联合分布能够确定边缘分布, 边缘分布却不能确定联合分布. 例如

$$p_1(x, y) = \begin{cases} x + y, & 0 \leq x, y \leq 1, \\ 0, & \text{其他}, \end{cases}$$

$$p_2(x, y) = \begin{cases} \left(\frac{1}{2} + x\right)\left(\frac{1}{2} + y\right), & 0 \leq x, y \leq 1, \\ 0, & \text{其他}, \end{cases}$$

均为联合密度函数, 显然 $p_1(x, y) \neq p_2(x, y)$, 但 $p_{1X}(x) = p_{2X}(x)$, $p_{1Y}(y) = p_{2Y}(y)$.

9. 相互独立的正态随机变量的线性组合是否仍为正态随机变量?

答 一般地, 有限个相互独立的正态随机变量的线性组合仍是正态随机变量, 如 $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, 可由卷积公式计算知 $Z = X_1 + X_2$ 仍服从正态分布, 而且 $Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, 即线性组合后仍然是正态随机变量, 其均值和方差是相应均值和方差的线性组合.

10. 协方差和相关系数反映随机变量 X 与 Y 之间的什么样的关系?

答 当给定两个随机变量 X 和 Y 时, 我们必然会考虑它们之间是否存在某种关系. 如当 X, Y 独立时, 有 $D(X + Y) = D(X) + D(Y)$; 一般情况下, $D(X + Y) = D(X) + D(Y) + 2E\{[X - E(X)][Y - E(Y)]\}$. 因而量 $E\{[X - E(X)][Y - E(Y)]\} \neq 0$ 反映了 X, Y 不独

立而存在某种相依关系的事实,将其定义为协方差.

当 $D(X), D(Y)$ 不变时, $\rho_{XY} = \text{cov}(X, Y) / [\sqrt{D(X)}\sqrt{D(Y)}]$ 反映了 X 和 Y 联系的密切程度. 当 $\rho_{XY} = \pm 1$ 时, X 和 Y 之间存在线性关系 $aX + bY + c = 0$ 的概率为 1; 而 $\rho_{XY} = 0$ 只反映 X, Y 不存在线性关系, 不排除其他的联系.

11. 两个随机变量相互独立与不相关有什么区别? 怎样解释它们之间的联系?

答 两个随机变量 X 与 Y 相互独立与不相关所反映的不是同一种关系. X 与 Y 的独立性反映 X 与 Y 之间不存在任何关系, 而 X 与 Y 不相关只是就线性关系而言的. 但当 (X, Y) 服从二维正态分布时, X 和 Y 独立与 X 和 Y 不相关是等价的. 详细的讨论可以得出如下结论.

(1) 若 X, Y 独立, 则 X, Y 不相关.

因为若 X, Y 独立, 则 $E(XY) = E(X)E(Y)$, 从而 $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 0$, 得 $\rho_{XY} = 0$, 所以 X, Y 不相关.

(2) 若 X, Y 不相关, 则 X, Y 不一定独立. 例如, 由 (X, Y) 的概率密度 $p(x, y) = \begin{cases} \frac{1}{\pi R^2}, & x^2 + y^2 \leq R^2 \\ 0, & \text{其他} \end{cases}$ 可算得 $E(X) = E(Y) = 0, E(X, Y) = 0$, 所以 $\rho_{XY} = 0$, X 与 Y 不相关. 但

$$p_X(x) = \frac{2\sqrt{1-x^2}}{\pi R^2}, |x| \leq R,$$

$$p_Y(y) = \frac{2\sqrt{1-y^2}}{\pi R^2}, |y| \leq R.$$

而 $p_X(x) \cdot p_Y(y) \neq p(x, y)$, 所以 X 与 Y 不独立.

(3) 若 X, Y 相关, 则 X 与 Y 不独立.

因为若 X, Y 相关, 则 $\rho_{XY} \neq 0$, 即 $\text{cov}(X, Y) \neq 0$, 因而 $E(XY) - E(X)E(Y) \neq 0$, 所以 X 与 Y 不独立.

(4) 若 X, Y 不独立, X 与 Y 不一定不相关.

12. 大数定律说明什么问题?

答 在实践中人们发现事件发生的“频率”具有稳定性, 在讨论数学期望时, 也看到在进行大量独立重复试验时, “平均值”也具有稳定性. 大数定律正是以严格的数学形式证明了“频率”或“平均值”的稳定性, 同时表达了这种稳定性的含义, 即“频率”或“平均值”在依概率收敛的意义下逼近某一常数.

13. 中心极限定理的意义是什么?

答 中心极限定理是阐明有些即使原来并不服从正态分布的一些独立的随机变量, 它们的总和的分布渐近地服从正态分布. 一般来说, 这些随机变量受到大量独立的因素中每项因素的影响是均匀的、微小的, 没有一项因素起特别突出的影响. 那么就可以断言, 这些随机变量的和的分布, 近似于正态分布.

14. 大数定律和中心极限定理之间, 有什么异同?

答 大数定律与中心极限定理都是通过极限理论来研究概率问题, 研究对象都是随机变量序列, 解决的都是概率论中的基本问题, 因而大数定律与中心极限定理在概率论中的意义十

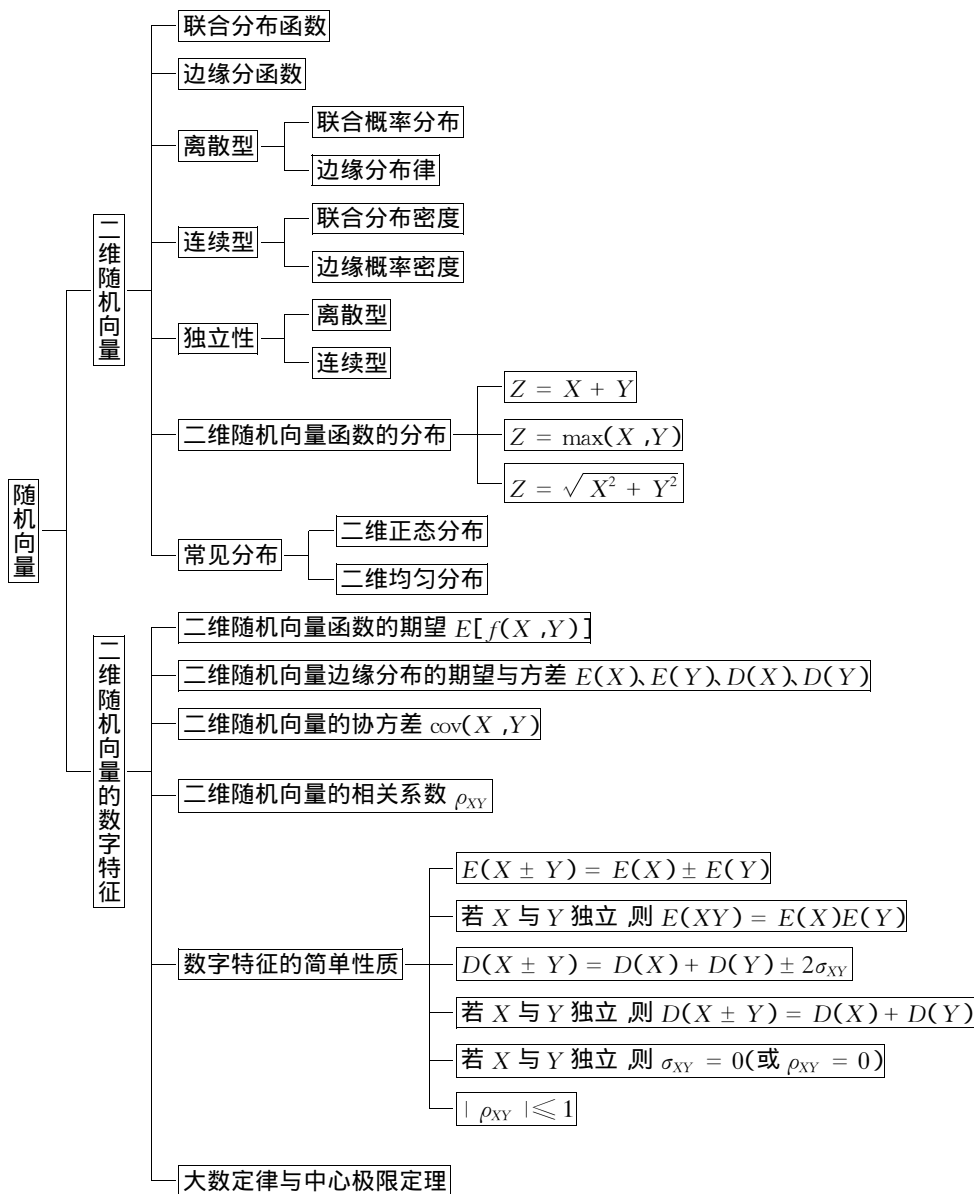
分重要.

它们的不同在于 :大数定律给出的是 ,当 $n \rightarrow \infty$ 时 ,随机变量序列的函数(平均值或频率)的极限 ,而中心极限定理则告诉我们 ,随机变量序列总和的分布近似正态分布 ,总和的标准化随机变量服从渐近标准正态分布 ,而不论随机变量序列服从何种分布 ,它是近两个世纪概率论研究的中心问题 ,所以称为中心极限定理.

本章常考题型归纳

考点	常见题型	例题
二维随机向量的联合分布函数与边缘分布函数 二维离散型随机向量的联合概率分布与边缘分布律 二维连续型随机向量的联合分布密度与边缘概率密度 独立性 二维随机向量函数的分布 二维正态分布与均匀分布 数字特征与性质 大数定律与中心极限定理	在实际问题中求二维离散型随机向量的联合分布	例 4.1.1
	已知分布或密度函数 ,确定函数中的参数 ,求分布函数或密度函数 ,并求事件的概率	例 4.1.2 例 4.1.3 例 4.2.4 例 4.2.5
	已知联合分布求边缘分布或反之	例 4.2.1 例 4.2.3 例 4.2.5
	判断 X 与 Y 的独立性或利用独立性求概率	例 4.2.1 例 4.2.2 例 4.2.6
	利用分布函数法求随随向量函数的分布	例 4.3.1 例 4.3.2 例 4.3.3
	利用公式求二维随机向量及随机向量函数的数字特征	例 4.4.1 例 4.4.2 例 4.4.4 例 4.4.5
	利用数字特征的性质进行证明	例 4.4.3
	利用 $E(X)$ 和 $D(X)$ 验证是否满足大数定律的条件	例 4.5.2
	在概率确定的条件下求样本数 m	例 4.5.3
	大数定律的应用	例 4.5.1

本章知识结构网络



检测题(一)

一、选择题

1. 设随机向量 (X, Y) 的密度函数为

$$p(x, y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{其他} \end{cases} \quad \text{则概率 } P(X < 0.5, Y < 0.6) \text{ 为 } (\quad).$$

A. 0.5

B. 0.3

C. $\frac{7}{8}$

D. 0.4

2. 设两个随机变量 X 与 Y 相互独立且同分布: $P(X = -1) = P(Y = -1) = \frac{1}{2}$, $P(X = 1) = P(Y = 1) = \frac{1}{2}$, 则下列各式中成立的是().

A. $P(X = Y) = \frac{1}{2}$ B. $P(X = Y) = 1$ C. $P(X + Y = 0) = \frac{1}{4}$ D. $P(XY = 1) = \frac{1}{4}$

3. 设两个相互独立的随机变量 X 和 Y 分别服从正态分布 $N(0, 1)$ 和 $N(1, 1)$, 则().

A. $P(X + Y \leq 0) = \frac{1}{2}$ B. $P(X + Y \leq 1) = \frac{1}{2}$ C. $P(X + Y \geq 0) = \frac{1}{2}$ D. $P(X - Y \leq 1) = \frac{1}{2}$

4. 设 X, Y 是相互独立的两个随机变量, 它们的分布函数分别为 $F_X(x), F_Y(y)$, 则 $Z = \max(X, Y)$ 的分布函数为().

A. $F_Z(z) = \max\{F_X(z), F_Y(z)\}$ B. $F_Z(z) = \max\{1 - F_X(z), 1 - F_Y(z)\}$

C. $F_Z(z) = F_X(z)F_Y(z)$ D. 都不是

5. 设随机变量 X 和 Y 相互独立, 其概率分布为

X	-1	1
p	$\frac{1}{2}$	$\frac{1}{2}$

,

Y	-1	1
p	$\frac{1}{2}$	$\frac{1}{2}$

,

则下列式子正确的是().

A. $X = Y$ B. $P(X = Y) = 0$ C. $P(X = Y) = \frac{1}{2}$ D. $P(X = Y) = 1$

6. 设两个相互独立的随机变量 X 和 Y 的方差分别为 4 和 2, 则随机变量 $3X - 2Y$ 的方差是().

A. 8 B. 16 C. 28 D. 44

7. 设随机变量 X 和 Y 的方差存在且不等于 0, 则 $D(X + Y) = D(X) + D(Y)$ 是 X 和 Y ().

A. 不相关的充分条件, 但不是必要条件 B. 独立的必要条件, 但不是充分条件

C. 不相关的充分必要条件 D. 独立的充分必要条件

8. 若随机变量 X 和 Y 的协方差 $\text{cov}(X, Y) = 0$, 则以下结论中正确的是().

A. X 与 Y 相互独立 B. $D(X + Y) = D(X) + D(Y)$

C. $D(X - Y) = D(X) - D(Y)$ D. $D(XY) = D(X)D(Y)$

二、填空题

1. 已知随机变量 X 和 Y 相互独立, 且都服从正态分布 $N(\mu, \sigma^2)$, 则 $E(X - Y) =$ _____, $D(X - Y) =$ _____.

2. 已知随机变量 $X \sim N(2, 4)$, $Y \sim N(3, 1)$, 且 X 与 Y 相互独立, 记 $Z = 2X - 3Y + 2$, 则 $Z \sim$ _____.

3. 随机向量 (X, Y) 在区域 $G: 0 \leq x \leq 1, x^2 \leq y \leq x$ 上服从均匀分布, 则其概率密度为 _____.

4. 设 (X, Y) 的联合分布律如下表所示, 则 $E(X, Y) =$ _____.

$\begin{array}{c} Y \\ \backslash X \end{array}$	0	1	2
-1	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{7}{20}$
2	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

5. 已知 $(X, Y) \sim p(x, y) = \begin{cases} C \sin(x+y), & 0 \leq x, y \leq \frac{\pi}{4} \\ 0, & \text{其他} \end{cases}$, 则 $C = \underline{\hspace{2cm}}$, Y 的边缘概

率密度 $p_Y(y) = \underline{\hspace{2cm}}$.

6. 设 $D(X) = 25$, $D(Y) = 36$, $\rho_{XY} = 0.4$, 则 $\text{cov}(X, Y) = \underline{\hspace{2cm}}$, $D(X+Y) = \underline{\hspace{2cm}}$, $D(X-Y) = \underline{\hspace{2cm}}$.

三、解答题

1. 将一枚均匀硬币掷三次, 以 X 记正面出现次数, 以 Y 记正面出现次数与反面出现次数之差的绝对值. 求随机向量 (X, Y) 的分布律.

2. 设随机向量 (X, Y) 的分布函数为

$$F(x, y) = \begin{cases} C - 3^{-x} - 3^{-y} + 3^{-x-y}, & x \geq 0, y \geq 0, \\ 0, & \text{其他}, \end{cases}$$

求: (1) 常数 C ; (2) 概率密度函数 $p(x, y)$.

3. 设随机向量 (X, Y) 的概率密度为

$$p(x, y) = \begin{cases} A(R - \sqrt{x^2 + y^2}), & x^2 + y^2 \leq R^2, \\ 0, & \text{其他} \end{cases}$$

求: (1) 系数 A 的值; (2) $P\{(X, Y) \in x^2 + y^2 \leq r^2\}$ ($r \leq R$).

4. 设二维随机向量 (X, Y) 的分布函数为

$$F(x, y) = A \left(B + \arctan \frac{x}{2} \right) \left(C + \arctan \frac{y}{3} \right), -\infty < x, y < +\infty.$$

求: (1) A, B, C 的值; (2) $p(x, y)$; (3) $p_X(x), p_Y(y)$.

5. 设随机向量 (X, Y) 的联合分布律为

$\begin{array}{c} X \\ \backslash Y \end{array}$	x_1	x_2	x_3
y_1	a	$\frac{1}{9}$	c
y_2	$\frac{1}{9}$	b	$\frac{1}{3}$

且 X 与 Y 相互独立, 试求 a, b, c 的值.

6. 设随机变量 X 和 Y 相互独立, 且

$$p_X(x) = \begin{cases} 1, & 0 \leq x \leq 1, \\ 0, & \text{其他}, \end{cases} \quad p_Y(y) = \begin{cases} 2y, & 0 \leq y \leq 1, \\ 0, & \text{其他}, \end{cases}$$

求随机变量 $Z = X + Y$ 的概率密度 $p_Z(z)$.

7. 已知 $X \sim N(1, 3^2)$, $Y \sim N(0, 4^2)$, $\rho_{XY} = -\frac{1}{2}$, 对 $Z = \frac{X}{3} + \frac{Y}{2}$, 求: (1) $E(Z)$ 和 $D(Z)$;

(2) ρ_{XZ} ; (3) X, Z 的独立性.

8. 设随机变量 $X \sim U[0, 2\pi]$ 概率密度

$$p_X(x) = \begin{cases} \frac{1}{2\pi}, & 0 \leq x \leq 2\pi, \\ 0, & \text{其他}, \end{cases}$$

令 $Y = \sin X$, $Z = \sin(X + a)$, $a \in [0, 2\pi]$ 为常数, 求: (1) ρ_{YZ} ; (2) 讨论 Y 与 Z 的独立性与相关性.

检测题(一)参考答案

一、选择题

1. B 2. A 3. B 4. C 5. C 6. D 7. B 8. B

二、填空题

1. $0.2\sigma^2$ 2. $N(-3, 25)$ 3. $p(x, y) = \begin{cases} 6, & (x, y) \in G, \\ 0, & \text{其他} \end{cases}$

4. $-\frac{3}{20}$ 5. $\sqrt{2} + 1$, $p_Y(y) = \begin{cases} -(\sqrt{2} + 1) \left[\cos\left(\frac{\pi}{4} + y\right) - \cos y \right], & 0 \leq y \leq \frac{\pi}{4}, \\ 0, & \text{其他} \end{cases}$

6. 12, 85, 37

三、解答题

		X			
		0	1	2	3
1.	Y = 1	0	$\frac{3}{8}$	$\frac{3}{8}$	0
	Y = 3	$\frac{3}{8}$	0	0	$\frac{1}{8}$

2. $C = 1$, $p(x, y) = \begin{cases} 3^{-x-y} (\ln 3)^2, & x \geq 0, y \geq 0, \\ 0, & \text{其他}. \end{cases}$

3. $A = \frac{3}{\pi R^3}$, $\frac{3r^2}{R^3} \left(1 - \frac{2r}{3R}\right)$.

4. $A = \frac{1}{\pi^2}$, $B = \frac{\pi}{2}$, $C = \frac{\pi}{2}$.

$$p(x, y) = \frac{6}{\pi^2(4+x^2)(9+y^2)}, -\infty < x, y < +\infty,$$

$$p_X(x) = \frac{2}{\pi(4+x^2)}, -\infty < x < +\infty,$$

$$p_Y(y) = \frac{3}{\pi(9+y^2)}, -\infty < y < +\infty.$$

5. $a = \frac{1}{18}$, $b = \frac{2}{9}$, $c = \frac{1}{6}$.

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

7. $E(Z) = \frac{1}{3}$, $D(Z) = 3$, $\rho_{XZ} = 0$, X, Z 相互独立.

$$8. \rho_{YZ} = \cos a.$$

当 $a = 0, \pi, 2\pi$ 时, $|\rho_{YZ}| = 1$, Y 与 Z 存在线性关系;

当 $a = \frac{\pi}{2}, \frac{3}{2}\pi$ 时, $\rho_{YZ} = 0$, Y 与 Z 不相关, 但 Y 与 Z 不独立.

检测题(二)

一、选择题

1. 设随机变量 (X, Y) 的概率密度为 $p(x, y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{其他} \end{cases}$, 则 $P(X < 0.5, Y < 0.5) = (\quad)$.

- A. 0.5 B. 0.3 C. $\frac{7}{8}$ D. 0.4

2. 下列四个二元函数, 哪个不能作为二维随机向量 (X, Y) 的分布函数().

$$A. F_1(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}), & 0 < x < +\infty, 0 < y < +\infty \\ 0, & \text{其他} \end{cases}$$

$$B. F_2(x, y) = \begin{cases} \sin x \sin y, & 0 \leq x \leq \frac{\pi}{2}, 0 \leq y \leq \frac{\pi}{2} \\ 0, & \text{其他} \end{cases}$$

$$C. F_3(x, y) = \begin{cases} 1, & x + 2y \geq 1 \\ 0, & x + 2y < 1 \end{cases}$$

$$D. F_4(x, y) = 1 + 2^{-x} - 2^{-y} + 2^{-x-y}$$

3. 设 X, Y 相互独立, 且都服从区间 $[0, 1]$ 上的均匀分布, 则服从区间或区域上的均匀分布的随机变量是().

- A. (X, Y) B. $X + Y$ C. X^2 D. $X - Y$

4. 设 $X_i \sim \frac{X_i}{p} \begin{vmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{vmatrix} (i = 1, 2)$, 且满足 $P(X_1 X_2 = 0) = 1$, 则 $P(X_1 = X_2) = (\quad)$.

- A. 0 B. $\frac{1}{4}$ C. $\frac{1}{2}$ D. 1

5. 已知随机变量 X 的方差 $D(X)$ 有限. 设 $Y = aX + b (a \neq 0, a, b \text{ 为常数})$, 则 $\rho_{XY} = (\quad)$.

- A. 1 B. -1 C. $\frac{a}{|a|}$ D. $|\rho_{XY}| < 1$

6. 若随机变量 X 与 Y 满足 $D(X + Y) = D(X - Y)$, 则必有().

- A. X 与 Y 相互独立 B. X 与 Y 不相关 C. $D(Y) = 0$ D. $D(X)D(Y) = 0$

7. 设随机向量 (X, Y) 服从二维正态分布, 则随机变量 $\zeta = X + Y$ 与 $\eta = X - Y$ 不相关的充分必要条件为().

- A. $E(X) = E(Y)$ B. $E(X^2) - [E(X)]^2 = E(Y^2) - [E(Y)]^2$
C. $E(X^2) = E(Y^2)$ D. $E(X^2) + [E(X)]^2 = E(Y^2) + [E(Y)]^2$

8. 将一枚硬币重复掷 n 次, 以 X 和 Y 表示正面向上和反面向上的次数, 则 X 和 Y 的相关系数等于().

- A. -1 B. 0 C. $\frac{1}{2}$ D. 1

二、填空题

1. 设离散型随机向量 (X, Y) 的分布为

$$p_{ij} = \frac{i+j}{30}, i=0, 1, 2, 3, j=0, 1, 2,$$

则 $P(X > 2, Y \leq 2) = \underline{\hspace{2cm}}, P(X + Y = 4) = \underline{\hspace{2cm}}.$

2. 设二维随机向量 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$, 其概率密度为 $\frac{1}{2\pi\sqrt{3}} \times e^{-\frac{4x^2 + 2xy + y^2 - 8x - 2y + 4}{6}}, -\infty < x, y < +\infty$, 则参数 $\mu_1 = \underline{\hspace{2cm}}, \mu_2 = \underline{\hspace{2cm}}, \sigma_1^2 = \underline{\hspace{2cm}}, \sigma_2^2 = \underline{\hspace{2cm}}, \rho = \underline{\hspace{2cm}}.$

3. 二维随机向量 (X, Y) 的概率密度为 $p(x, y) = \begin{cases} \frac{1}{4}, & |x| < 1, |y| < 1 \\ 0, & \text{其他} \end{cases}$, 则关于 X 与 Y

的边缘分布 $p_X(x) = \underline{\hspace{2cm}}, p_Y(y) = \underline{\hspace{2cm}}.$

4. 掷 2 枚骰子, 以 X 记第一枚骰子掷出的点数, 以 Y 记第二枚骰子掷出的点数, 则 $E(X + Y) = \underline{\hspace{2cm}}, E(XY) = \underline{\hspace{2cm}}.$

5. 设对随机变量 X, Y 有 $E(X) = 2, E(X^2) = 20, E(Y) = 3, E(Y^2) = 34, \rho_{XY} = 0.5$, 则 $E(3X + 2Y) = \underline{\hspace{2cm}}, D(X - Y) = \underline{\hspace{2cm}}.$

6. 设 X 和 Y 是两个相互独立且服从正态分布 $N\left(0, \left(\frac{1}{\sqrt{2}}\right)^2\right)$ 的随机变量, 则随机变量 $|X - Y|$ 的数学期望 $E|X - Y| = \underline{\hspace{2cm}}.$

三、解答题

1. 设 ξ, η 是相互独立且服从同一分布的两个离散型随机变量, 已知 ξ 的分布律为 $P(\xi = i) = \frac{1}{3}, i = 1, 2, 3$, 又设 $X = \max(\xi, \eta), Y = \min(\xi, \eta)$, 求 (X, Y) 的联合分布律和边缘分布律.

2. 设 (X, Y) 在区域 $D = \{(x, y) | 0 < x < 2, -1 < y < 2\}$ 上服从均匀分布, 试求 $P(X \leq Y), P(X + Y > 1).$

3. 设随机向量 (X, Y) 的分布函数 $F(x, y) = A\left(B + \arctan \frac{x}{2}\right)\left(C + \arctan \frac{y}{2}\right)$. 求: (1) 系数 A, B, C ; (2) (X, Y) 的分布密度; (3) 边缘分布密度. 随机变量 X 与 Y 是否独立?

4. 设随机向量 (X, Y) 的分布密度为 $\varphi(x, y) = \frac{A}{(x^2 + y^2 + 1)^2}$. 求: (1) 系数 A ; (2) $E(X), E(Y), D(X), D(Y).$

5. 设随机向量 (X, Y) 在区域 $D = \{(x, y) | 0 < x < 1, |y| < x\}$ 内服从均匀分布, 求关于 X 的边缘概率密度函数及随机变量 $Z = 2X + 1$ 的方差 $D(Z).$

6. 设 $W = (aX + 3Y)^2, E(X) = E(Y) = 0, D(X) = 4, D(Y) = 16, \rho_{XY} = -0.5$, 求常数 a

使 $E(W)$ 最小, 并求 $E(W)$ 的最小值.

检测题(二)参考答案

一、选择题

1. B 2. C 3. A 4. A 5. C 6. B 7. B 8. A

二、填空题

1. $\frac{2}{5}, \frac{4}{15}$ 2. 1 0 1 4, $-\frac{1}{2}$

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

4. 7 12.25 5. 12 21 6. $\sqrt{\frac{2}{\pi}}$

三、解答题

		X			$p_{\cdot j}$	
		1	2	3		
1.	Y					
	1	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{5}{9}$	
	2	0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	
	3	0	0	$\frac{1}{9}$	$\frac{1}{9}$	
		$p_{i \cdot}$	$\frac{1}{9}$	$\frac{3}{9}$	$\frac{5}{9}$	1

2. $\frac{1}{3}, \frac{2}{3}$.

3. $A = \frac{1}{\pi^2}, B = C = \frac{\pi}{2}, p(x, y) = \frac{6}{\pi^2(4+x^2)(9+y^2)},$

$p_X(x) = \frac{2}{\pi(4+x^2)}, p_Y(y) = \frac{3}{\pi(9+y^2)}, X$ 与 Y 相互独立.

4. $A = \frac{1}{\pi}, E(X) = 0, E(Y) = 0, D(X) = D(Y) = \infty.$

5. $p_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{其他}, \end{cases} D(Z) = \frac{2}{9}.$

6. 当 $a = 3$ 时, $E(W)_{\min} = 108.$

第5章 样本及抽样分布

5.1 总体与样本

5.1.1 知识要点

1. 定义

总体与样本 研究对象的某项数量指标的全体称为总体.

在一个总体 X 中, 抽取 n 个个体 $X_1, X_2, \dots, X_n$, 这 n 个个体 $X_1, X_2, \dots, X_n$ 称为总体 X 的一个容量为 n 的样本(或称子样).

简单随机样本 设 X 是一个具有分布函数 F 的随机变量, 从总体 X 中抽取的有同一分布 F 的 n 个相互独立的随机变量 $X_1, X_2, \dots, X_n$ 称为总体 X 的一个容量为 n 的简单随机样本.

2. 分布定理

若 $X_1, X_2, \dots, X_n$ 是来自总体 X 的一个样本, $F(x)$ 是 X 的分布函数, $p(x)$ 是 X 的密度函数, 则 $X_1, X_2, \dots, X_n$ 的联合分布函数 $F^*(x_1, x_2, \dots, x_n)$ 和联合概率密度 $p^*(x_1, x_2, \dots, x_n)$ 分别为

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

$$p^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i).$$

5.1.2 典型题型解析

例1 设总体 $X \sim P\{X=m\} = q^{m-1}p, m=1, 2, 3, \dots$, 其中 $0 < p < 1 (q=1-p)$ 为参数, 试求样本 $X_1, X_2, \dots, X_n$ 的联合分布列.

解 设 $x_1, x_2, \dots, x_n$ 是样本 $X_1, X_2, \dots, X_n$ 的一组观察值, 于是样本的联合分布列为

$$P\{X_1=x_1, X_2=x_2, \dots, X_n=x_n\} = \prod_{i=1}^n P\{X_i=x_i\} = \prod_{i=1}^n q^{x_i-1}p = p^n q^{(\sum_{i=1}^n x_i)-n}.$$

注 样本的联合分布列(密度)是 n 维随机变量相互独立的充要条件的体现. 具体表达式总是建立在某一样本点上, 因而是样本点的函数.

为处理极大似然估计问题的需要, 给出联合分布列(密度)时, 只要有可能, 尽量把“ $\prod_{i=1}^n \dots$ ”

转化为“ $\sum_{i=1}^n$ ”表出, 否则, 仍保留“ $\prod_{i=1}^n \dots$ ”.

例2 设总体 X 有分布密度

$$p(x) = \begin{cases} \frac{4x^2}{a^2\sqrt{\pi}} e^{-\frac{x^2}{a^2}}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad a > 0 \text{ 为参数.}$$

试求样本 $X_1, X_2, \dots, X_n$ 的联合密度.

解 设 $X_1, X_2, \dots, X_n$ 是样本点, 于是在 $x_i > 0, i=1, 2, \dots, n$ 的公共区域上, 样本的联合密度为

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P_{X_i}(x_i) = \prod_{i=1}^n \frac{4x_i^2}{a^2\sqrt{\pi}} e^{-\frac{x_i^2}{a^2}} = \left(\frac{4}{\sqrt{\pi}}\right)^n \frac{1}{a^{3n}} \left(\prod_{i=1}^n x_i\right)^2 e^{-\frac{1}{a^2} \sum_{i=1}^n x_i^2}.$$

否则 $p(x_1, x_2, \dots, x_n) = 0$.

5.2 抽样分布

5.2.1 知识要点

1. 定义

统计量: 设 $X_1, X_2, \dots, X_n$ 是来自总体 X 的一个样本, $g(X_1, X_2, \dots, X_n)$ 是 $X_1, X_2, \dots, X_n$ 的函数, 且 g 中不含任何未知参数, 则称 $g(X_1, X_2, \dots, X_n)$ 是一统计量.

观察值: 设 $x_1, x_2, \dots, x_n$ 是样本 $X_1, X_2, \dots, X_n$ 的样本值, 则称 $g(x_1, x_2, \dots, x_n)$ 是 $g(X_1, X_2, \dots, X_n)$ 的观察值.

2. 常用统计量

$$\text{样本均值 } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

$$\text{样本方差 } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right],$$

$$\text{样本标准差 } S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2},$$

$$\text{样本 } k \text{ 阶原点矩 } A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k=1, 2, \dots,$$

$$\text{样本 } k \text{ 阶中心矩 } B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k=1, 2, \dots.$$

它们的观察值分别为

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right],$$

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

$$a_k = \frac{1}{n} \sum_{i=1}^n x_i^k, k=1, 2, \dots,$$

$$b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, k=1, 2, \dots.$$

这些观察值分别称为样本均值、样本方差、样本标准差、样本 k 阶原点矩、样本 k 阶中心矩.

3. 统计量的分布

(1) 样本均值 $\bar{X}$ 的分布

设 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \dots, X_n$ 是 X 的一个样本. 则样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 是统计量, 且

$$\bar{X} \sim N(\mu, \sigma^2/n) \text{ 或 } (\bar{X} - \mu) / \sqrt{\sigma^2/n} \sim N(0, 1),$$

$(\bar{X} - \mu) / \sqrt{\sigma^2/n}$ 也称 U 统计量.

(2) χ^2 分布

设 $X \sim N(0, 1)$, $X_1, X_2, \dots, X_n$ 是 X 的一个样本, 则统计量

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

的分布称为服从自由度为 n 的 χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$, $\chi^2(n)$ 分布的概率密度函数为

$$f(y) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} e^{-y/2}, & y > 0, \\ 0, & \text{其他.} \end{cases}$$

若 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \dots, X_n$ 是 X 的一个样本, 则统计量

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n).$$

χ^2 分布具有以下性质:

1° 若 $\chi^2 \sim \chi^2(n)$, 则 $E(\chi^2) = n$, $D(\chi^2) = 2n$.

2° 若 $X_1 = \chi_1^2 \sim \chi^2(n_1)$, $X_2 = \chi_2^2 \sim \chi^2(n_2)$, 且 X_1, X_2 相互独立, 则

$$X_1 + X_2 = \chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$$

此结果可以推广到有限个 χ^2 分布相加的情形.

(3) t 分布

设总体 $X \sim N(0, 1)$, $Y \sim \chi^2(n)$, 且 X, Y 相互独立, 则称随机变量

$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布, 记作 $t \sim t(n)$, $t(n)$ 分布的概率密度函数为

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < t < +\infty.$$

t 分布具有以下性质:

$$\lim_{n \rightarrow \infty} f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

对 $t(n)$ 的上 α 分位点, 由 $f(t)$ 的图形对称性得

$$t_{1-\alpha}(n) = -t_{\alpha}(n).$$

(4) F 分布

设总体 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, U 与 V 相互独立, 则称统计量 $F = \frac{U/n_1}{V/n_2}$ 服从第一自由度为 n_1 , 第二自由度为 n_2 的 F 分布, 记作 $F \sim F(n_1, n_2)$, $F(n_1, n_2)$ 的概率密度为

$$f(y) = \begin{cases} \frac{\Gamma[(n_1 + n_2)/2] \Gamma(n_1/2) \Gamma(n_2/2) y^{(n_1/2)-1}}{\Gamma(n_1/2) \Gamma(n_2/2) [1 + (n_1 y / n_2)]^{(n_1 + n_2)/2}}, & y > 0, \\ 0, & \text{其他.} \end{cases}$$

若 $F \sim F(n_1, n_2)$, 则 $\frac{1}{F} \sim F(n_2, n_1)$.

(5) 正态总体样本方差 S^2 的分布

设总体 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \dots, X_n$ 是 X 的一个样本, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 是样本方差, 则① $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, ② $\bar{X}$ 与 S^2 相互独立.

由此可以推出

①若 $X_1, X_2, \dots, X_n$ 是 $X \sim N(\mu, \sigma^2)$ 的一个样本, 则

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

②若 $X_1, X_2, \dots, X_{n_1}$ 是 $X \sim N(\mu, \sigma^2)$ 的一个样本, $Y_1, Y_2, \dots, Y_{n_2}$ 是 $Y \sim N(\mu_2, \sigma^2)$ 的一个样本, X, Y 相互独立, $\bar{X}, \bar{Y}$ 是两样本的样本均值, S_1^2, S_2^2 是两样本的样本方差, 则

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

其中 $S_w^2 = [(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2] / (n_1 + n_2 - 2)$.

③设 $X_1, X_2, \dots, X_n$ 是 $X \sim N(\mu_1, \sigma_1^2)$ 的一个样本, $Y_1, Y_2, \dots, Y_{n_2}$ 是 $Y \sim N(\mu_2, \sigma_2^2)$ 的一个样本, 它们相互独立. 若 S_1^2, S_2^2 是它们的样本方差, 则

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

5.2.2 典型题型解析

1. 关于统计量的定义

例 5.2.1 样本 X_1, X_2, X_3, X_4 取自正态分布总体 X , $EX = \mu$ 为已知, 而 $DX = \sigma^2$ 未知, 则下列随机变量中不能作为统计量的是().

A. $\bar{X} = \frac{1}{4} \sum_{i=1}^n X_i$

B. $X_1 + X_4 - 2\mu$

C. $K = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$

D. $S^2 = \frac{1}{3} \sum_{i=1}^n (X_i - \bar{X})^2$

解 统计量不仅是样本的函数, 而且其中不能包含总体分布中的未知参数, 尽管 B 中含

有参数 μ , 但因 μ 已知, 它仍可以作为统计量, 而对于 C 则因包含总体分布中的未知参数 σ^2 , 不能作为统计量. 综上所述, 应选择 C.

例 5.2.2 设总体 $X \sim N(\mu, \sigma^2)$, 其中 μ 是已知的, 而 σ^2 是未知的, X_1, X_2, X_3 是总体的一个样本, 试问 $X_1 + X_2 + X_3, X_1 + 2\mu, \max(X_1, X_2, X_3)$ 及 $\sum_{i=1}^3 \frac{X_i^2}{\sigma^2}, \frac{X_1 + X_2}{2}$ 之中, 哪些是统计量? 哪些不是统计量?

解 $X_1 + X_2 + X_3, \max(X_1, X_2, X_3), \frac{X_1 + X_2}{2}$ 均为样本函数, 且不含未知数, 所以都是统计量.

$X_1 + 2\mu$ 虽然含有参数 μ , 但 μ 是已知的参数, 所以 $X_1 + 2\mu$ 也是统计量. 而 $\sum_{i=1}^3 \frac{X_i^2}{\sigma^2}$ 虽为样本函数, 但含未知数 σ^2 , 所以 $\sum_{i=1}^3 \frac{X_i^2}{\sigma^2}$ 不是统计量.

注 样本是总体的反映, 又是进行统计推断的依据, 但样本反映的信息是零乱的、无序的和分散的. 所以要针对不同的问题构造样本的不同函数, 将信息集中起来, 便于进行统计推断和研究分析, 使之更易揭示问题的本质. 统计量就是样本的不含未知参数的连续函数. 例如 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 是一个统计量, 不含任何未知参数, 它排除了关于 σ^2 的信息, 集中了关于 μ 的信息.

连续性是为了保证统计量仍然是随机变量而提出的, 同时也为了以后用数学分析方法研究统计量(如极大似然函数)而提供了条件.

不含未知参数, 则统计量只与样本有关, 而与总体无关. 若含有未知参数, 则无法依靠样本规则值求未知参数的估计值, 因而失去利用统计量估计未知参数的作用. 这就违背了引进统计量的初衷.

2. 利用统计量的分布确定未知参数

例 5.2.3 设 X_1, X_2, X_3, X_4 是来自总体 $X \sim N(0, 1)$ 的一个样本, 问: a, b 取何值时, $Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2 \sim \chi^2(n)$? 并确定 n 的值.

解 因为 X_1, X_2, X_3, X_4 独立同分布, 所以

$$E(X_1 - 2X_2) = 0, E(3X_3 - 4X_4) = 0,$$

$$D(X_1 - 2X_2) = D(X_1) + 4D(X_2) = 20,$$

$$D(3X_3 - 4X_4) = 9D(X_3) + 16D(X_4) = 100,$$

于是 $\frac{X_1 - 2X_2}{\sqrt{20}} \sim N(0, 1), \frac{3X_3 - 4X_4}{10} \sim N(0, 1),$

而且 $X_1 - 2X_2$ 与 $3X_3 - 4X_4$ 相互独立, 此时

$$\frac{(X_1 - 2X_2)^2}{20} + \frac{(3X_3 - 4X_4)^2}{100} \sim \chi^2(2).$$

从而 $a = \frac{1}{20}, b = \frac{1}{100}, n = 2.$

例 5.2.4 设总体 $X \sim N(0, 1), X_1, X_2, \dots, X_6$ 为 X 的一个样本, 令

$$Y = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2.$$

求常数 c , 使 $cY \sim \chi^2$ 分布.

解 因为各 X_i 独立同分布, 所以

$$X_1 + X_2 + X_3 \sim N(0, 3), X_4 + X_5 + X_6 \sim N(0, 3),$$

$$(X_1 + X_2 + X_3)/\sqrt{3} \sim N(0, 1),$$

$$(X_4 + X_5 + X_6)/\sqrt{3} \sim N(0, 1).$$

于是

$$\begin{aligned} & (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2 \\ &= \frac{1}{3} \left[\left(\frac{X_1 + X_2 + X_3}{\sqrt{3}} \right)^2 + \left(\frac{X_4 + X_5 + X_6}{\sqrt{3}} \right)^2 \right] \\ &= Y_1^2 + Y_2^2, \end{aligned}$$

其中 $Y_1^2/3 \sim \chi^2(1), Y_2^2/3 \sim \chi^2(1),$

所以 $\frac{1}{3}(Y_1^2 + Y_2^2) = \frac{1}{3}[(X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2] = \frac{1}{3}Y \sim \chi^2(2).$

即 $c = \frac{1}{3}.$

3. 确定统计量的分布

例 5.2.5 设 $X_1, X_2, \dots, X_{n_1}$ 是总体 $X \sim N(\mu_1, \sigma_1^2)$ 的一个样本, $Y_1, Y_2, \dots, Y_{n_2}$ 是总体 $Y \sim N(\mu_2, \sigma_2^2)$ 的一个样本, 且两个样本相互独立, 令

$$\hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_1)^2, \hat{\sigma}_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - \mu_2)^2.$$

求 $F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ 的抽样分布 (μ_1, μ_2 已知)

解 因为

$$\chi_1^2 = \frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} (X_i - \mu_1)^2 \sim \chi^2(n_1),$$

$$\chi_2^2 = \frac{1}{\sigma_2^2} \sum_{i=1}^{n_2} (Y_i - \mu_2)^2 \sim \chi^2(n_2).$$

χ_1^2 与 χ_2^2 相互独立, 由 F 分布定义知

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F(n_1, n_2).$$

因为我们一般只习惯于总体 $X \sim N(0, 1)$ 下的 χ^2 分布, 而不习惯总体 $X \sim N(\mu, \sigma^2)$ 下的 χ^2 分布, 所以必须适应这种转换.

例 5.2.6 设 $X_1, X_2, \dots, X_n$ 是总体 $X \sim N(\mu, \sigma^2)$ 的一个样本, μ, σ^2 均未知, 则下列结论 () 正确.

A. $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$ B. $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$

$$C. \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$$

$$D. \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n)$$

解 选 C. 当 $X \sim N(\mu, \sigma^2)$ 时, 有 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$, 但 μ, σ^2 未知. 以 $\bar{X}$ 代替 μ , 则 $\bar{X}$ 为一个约束条件, 所以 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$ 故选 C.

例 5.2.7 设 $X_1, X_2, \dots, X_5$ 是总体 $X \sim N(0, 1)$ 的一个样本, 若统计量

$$U = c(X_1 + X_2) / \sqrt{X_3^2 + X_4^2 + X_5^2} \sim t(n).$$

试确定 c 与 n .

解 因为 X_i 独立同分布 ($i=1, 2, \dots, 5$), 所以 $(X_1 + X_2) / \sqrt{2} \sim N(0, 1)$, $X_3^2 + X_4^2 + X_5^2 \sim \chi^2(3)$. 且两者相互独立, 由 t 分布定义知

$$U = \frac{X_1 + X_2}{\sqrt{2}} / \sqrt{(X_3^2 + X_4^2 + X_5^2)/3} \sim t(3),$$

故 $c = \sqrt{\frac{3}{2}}, n = 3$.

例 5.2.8 设 X_1, X_2 为总体 $X \sim N(0, \sigma^2)$ 的一个样本, 问: $Y = (X_1 + X_2)^2 / (X_1 - X_2)^2$ 服从什么分布? 参数为何?

解 因为 $X \sim N(0, \sigma^2)$, X_1, X_2 相互独立, 所以 $(X_1 + X_2)$ 和 $(X_1 - X_2)$ 都服从正态分布 $N(0, 2\sigma^2)$, 且

$$[(X_1 + X_2) / (\sqrt{2}\sigma)]^2 \sim \chi^2(1), [(X_1 - X_2) / (\sqrt{2}\sigma)]^2 \sim \chi^2(1).$$

从而由 F 分布定义知

$$Y = \frac{[(X_1 + X_2) / (\sqrt{2}\sigma)]^2}{[(X_1 - X_2) / (\sqrt{2}\sigma)]^2} = \frac{(X_1 + X_2)^2}{(X_1 - X_2)^2} \sim F(1, 1),$$

即 Y 服从 F 分布, 参数为 $(1, 1)$.

同步练习 5.2

1. 设容量 $n = 10$ 的样本观察值为 $(8, 7, 6, 5, 9, 8, 7, 5, 9, 6)$. 求样本均值及样本方差的观测值.

2. 假设随机变量 X 服从正态分布 $N(1, 2^2)$, $X_1, \dots, X_{100}$ 是来自 X 的样本, $\bar{X}$ 为样本均值. 已知 $Y = a\bar{X} + b \sim N(0, 1)$, 求 a, b 的值.

3. 设 $X_1, X_2, \dots, X_n$ 为来自正态总体 $X \sim N(\mu, \sigma^2)$ 的一个简单随机样本, 则样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 服从什么分布; 若 a_i 为常数 ($a_i \neq 0, i = 1, 2, \dots, n$), 则 $\sum_{i=1}^n a_i X_i$ 服从什么分布?

4. 已知样本 $X_1, \dots, X_{10}$ 取自正态总体分布 $N(0, 1)$, $\bar{X}$ 为样本均值, 已知 $P(\bar{X} \geq \lambda) = 0.01$, 则 λ 的值为多少?

5. 假设 $X \sim N(0, 1)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, 服从自由度为 $(n-1)$ 的 χ^2 分布的随机变量是().

A. $\sum_{i=1}^n X_i^2$ B. S^2 C. $(n-1)\bar{X}^2$ D. $(n-1)S^2$

6. 设 $X_1, X_2, \dots, X_8$ 和 $Y_1, Y_2, \dots, Y_{10}$ 是分别来自两个正态总体 $N(-1, 2)$ 和 $N(2, 5)$ 的样本且相互独立, S_1^2 和 S_2^2 分别为两个样本的样本方差. 则服从 $F(7, 9)$ 的统计量为().

A. $\frac{2S_1^2}{5S_2^2}$ B. $\frac{5S_1^2}{4S_2^2}$ C. $\frac{4S_2^2}{5S_1^2}$ D. $\frac{5S_1^2}{2S_2^2}$

同步练习 5.2 参考答案

1. 7.2.

2. -5.5.

3. $N\left(\mu, \frac{\sigma^2}{n}\right), N\left(\sum_{i=1}^n a_i \mu, \sigma^2 \sum_{i=1}^n a_i^2\right).$

4. 0.58.

5. D

6. D

5.3 条形图、直方图及经验分布函数

5.3.1 知识要点

1. 频率直方图

设 $x_1, x_2, \dots, x_n$ 是取自总体 X 的一个样本的样本值, 作频率直方图步骤如下.

(1) 找出数据 $x_1, x_2, \dots, x_n$ 中的最小值和最大值, 记 $x_1^* = \min\{x_1, x_2, \dots, x_n\}$, $x_n^* = \max\{x_1, x_2, \dots, x_n\}$.

(2) 选 a (它略小于 x_1^*) b (它略大于 x_n^*), 并等分区间 $(a, b]$, 得

$$a = t_0 < t_1 < t_2 < \dots < t_{m-1} < t_m = b.$$

于是得 m 个小区间

$$(t_0, t_1], (t_1, t_2], \dots, (t_{m-1}, t_m],$$

其中组距 $\Delta t_i = t_i - t_{i-1} = \frac{b-a}{m}$ ($i=1, 2, \dots, m$). 子区间的个数 m 一般没有硬性规定, 当样本容量 n 小时, m 也应小些, n 大时, m 则大些. 在一般情况下可取 8 到 15. 另外, 分点应比数据的有效数字多一位.

(3) 计算数据落入各区间的频数 n_i 及频率 $f_i = \frac{n_i}{n}$ ($i=1, 2, \dots, m$).

(4) 在横轴截取各子区间, 并以各子区间为底, 以 $f_i / \Delta t_i$ 为高做出 n 个小矩形, 每个小矩形的面积 Δs_i 就等于数据落入第 i 个小区间的频率, 即 $\Delta s_i = \Delta t_i \cdot f_i / \Delta t_i = f_i$, 于是, 有

$$\sum_{i=1}^n \Delta s_i = \sum_{i=1}^m f_i = \sum_{i=1}^m n_i / n = 1.$$

这样就完成了频率直方图.

2. 经验分布函数

设 $x_1, x_2, \dots, x_n$ 是总体 X 的一组样本值, 把它们从小到大按顺序排列, 得到

$$x_1^* \leq x_2^* \leq \dots \leq x_n^*.$$

这里 x_1^* 是 $x_1, x_2, \dots, x_n$ 中最小的数, x_i^* 是 $x_1, x_2, \dots, x_n$ 中第 i 个小的数, x_n^* 是 $x_1, x_2, \dots, x_n$ 中最大的数, 记

$$F_n(x) = \begin{cases} 0, & x < x_1^* \\ \frac{k}{n}, & x_k^* \leq x < x_{k+1}^*, \quad k = 1, 2, \dots, n-1, \\ 1, & x \geq x_n^*. \end{cases}$$

$F_n(x)$ 是一非减连续函数, 且满足

$$F_n(-\infty) = 0, F_n(+\infty) = 1.$$

称 $F_n(x)$ 为总体 X 的经验分布函数(或称子样分布函数).

3. 格利文科定理

分布函数 $F_n(x)$ 依概率 1 关于 x 一致收敛于总体分布函数 $F(x)$, 即

$$P\left\{\lim_{n \rightarrow \infty} \sup_{-\infty < x < +\infty} |F_n(x) - F(x)| = 0\right\} = 1.$$

5.3.2 典型题型解析

例 5.3.1 设 X 的子样 $(X_1, X_2, \dots, X_9)$ 的一个观察值为 $(1.3, 2.5, 2.4, 2.5, 3, 2.5, 5)$. 求 X 的经验分布函数 $F_n(x)$.

解 由 $F_n(x)$ 的定义得

$$F_9(x) = \begin{cases} 0, & x < 1, \\ \frac{1}{9}, & 1 \leq x < 2, \\ \frac{2}{9}, & 2 \leq x < 2.5, \\ \frac{5}{9}, & 2.5 \leq x < 3, \\ \frac{7}{9}, & 3 \leq x < 4, \\ \frac{8}{9}, & 4 < x \leq 5, \\ 1, & x > 5. \end{cases}$$

疑难解析

1. 设 $X_1, X_2, \dots, X_n$ 为总体 X 的样本, 称 $X_1, X_2, \dots, X_n$ 为简单随机样本, 其含义是什么? 简述对样本如此处理的必要性.

答 把 $X_1, X_2, \dots, X_n$ 看作简单随机样本, 严格地说, 这是一种约定, 而这个约定的含义

是由独立性、代表性所阐述的.

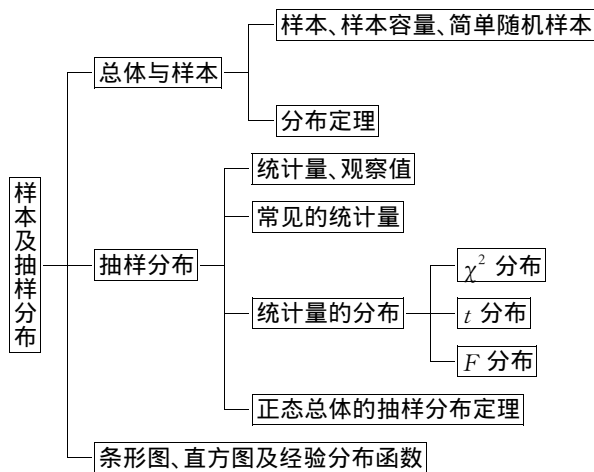
在上述约定下,概率论中有关独立条件下的一系列结论,都可以引入到数理统计中来,成为统计推断不可缺少的工具.

分析:简单随机样本依托的实际模型是放回抽样.如此抽样的特点是每次抽得的结果既不影响其他批次抽样结果,也不受其他批次抽样结果的影响.不放回抽样产生的样本,由于既不满足独立性又不满足代表性,因而不能成为简单随机样本.

本章常考题型归纳

考 点	常见题型	例 题
统计量的定义	确定统计量	例 5.2.1 例 5.2.2
统计量的分布	利用统计量的分布确定参数	例 5.2.3 例 5.2.4
	确定统计量的分布	例 5.2.5 例 5.2.6 例 5.2.7 例 5.2.8

本章知识结构网络



检测题(一)

一、填空题

1. 设 $X_1, X_2, \dots, X_n$ 是来自指数分布 $E(\lambda)$ 的简单随机样本, $\lambda > 0$ 为未知参数, 则 $(X_1, X_2, \dots, X_n)$ 的概率分布为 _____; 设 $n = 10$ 时, 样本的一组观测值为 $(4.6, 4.3, 5.4, 5.8, 4.7)$, 则样本均值为 _____, 样本方差为 _____.

2. 已知样本 $X_1, X_2, \dots, X_{16}$ 取自正态分布总体 $N(0, 1)$, $\bar{X}$ 为样本均值, 已知 $P\{\bar{X} \geq \lambda\} = 0.01$, 则 $\lambda =$ _____.

3. 设随机变量 X 服从正态分布 $N(\mu, \sigma^2)$, $X_1, \dots, X_{20}$ 是来自 X 的一个样本, 令

$$Y = 3 \sum_{i=1}^{10} X_i - 4 \sum_{i=11}^{20} X_i$$

则 Y 服从分布 _____.

4. 假设随机变量 X 服从正态分布 $N(0, 1)$. 若 $P\{2X^2 \geq \lambda\} = 0.05$, 则 $\lambda =$ _____.

5. 设总体 $X \sim N(\mu, \sigma^2)$, $\bar{X}$ 是样本均值, S_n^2 是样本方差, n 为样本容量, 则常用统计量 $U =$ _____ $\sim N(0, 1)$, $T =$ _____ $\sim t(n-1)$; $\chi^2 =$ _____ $\sim \chi^2(n-1)$.

二、选择题

1. 设随机变量 X 服从正态分布 $N(1, 2^2)$, $X_1, \dots, X_{100}$ 是来自 X 的样本, $\bar{X}$ 为样本均值, 已知 $Y = a\bar{X} + b \sim N(0, 1)$, 则有

A. $a = -5, b = 5$

B. $a = 5, b = 5$

C. $a = \frac{1}{5}, b = -\frac{1}{5}$

D. $a = -\frac{1}{5}, b = \frac{1}{5}$

2. 样本 $X_1, \dots, X_n$ 来自正态分布总体 $N(\mu, \sigma^2)$, $\bar{X}$ 与 S^2 分别是样本均值和样本方差, 则下面结论不成立的有().

A. $\bar{X}$ 与 S^2 相互独立

B. $\bar{X}$ 与 $(n-1)S^2$ 相互独立

C. $\bar{X}$ 与 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$ 相互独立

D. $\bar{X}$ 与 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2$ 相互独立

3. 样本 $(X_1, X_2, \dots, X_n)$ 取自标准正态分布总体 $N(0, 1)$, $\bar{X}, S$ 分别为样本均值及标准差, 则().

A. $\bar{X} \sim N(0, 1)$

B. $\sqrt{n}\bar{X} \sim N(0, 1)$

C. $\sum_{i=1}^n X_i^2 \sim \chi^2(n-1)$

D. $\bar{X}/S \sim t(n-1)$

4. 假设 $X \sim N(0, 1)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, 服从自由度为 $(n-1)$ 的 χ^2 分布的随机变量是().

A. $\sum_{i=1}^n X_i^2$

B. S^2

C. $(n-1)\bar{X}^2$

D. $(n-1)S^2$

5. 设 $X \sim N(0, \sigma^2)$, 则服从自由度为 $n-1$ 的 t 分布的随机变量是().

A. $\frac{\sqrt{n}\bar{X}}{S}$

B. $\frac{\sqrt{n-1}\bar{X}}{S}$

C. $\frac{\sqrt{n}\bar{X}}{S^2}$

D. $\frac{\sqrt{n-1}\bar{X}}{S^2}$

三、解答题

1. 自总体 X 抽得一个容量为 5 的样本 8, 2, 5, 3, 7. 求样本均值 $\bar{X}$, 样本方差 S^2 , 经验分布函数 $F_n(x)$.

2. 设总体 X 服从参数为 λ 的指数分布, 样本 $(X_1, X_2, \dots, X_n)$ 来自总体区. 求样本均值与样本方差的均值.

3. 假设总体 X 的密度为

$$f(x, \mu) = \begin{cases} (\theta+1)x^\theta, & 0 \leq x \leq 1, \\ 0, & \text{其他}, \end{cases}$$

其中 $\theta > 0$. 求来自 X 的简单随机样本 $(X_1, \dots, X_n)$ 的密度 $f(x_1, \dots, x_n)$.

4. 假设随机变量 X 服从分布 $\chi^2(5)$. 求 λ 的值使其满足

$$P\{X \leq \lambda\} = 0.05.$$

5. 设 $X_1, X_2, \dots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的一个样本, S^2 为样本方差. 求满足不等

式 $P\left\{\frac{S^2}{\sigma^2} \leq 1.5\right\} \geq 0.95$ 的最小 n 值.

6. 设 X_1, X_2 为总体 $X \sim N(0, \sigma^2)$ 的一个样本, 则 $Y = (X_1 + X_2)^2 / (X_1 - X_2)^2$ 服从什么分布?

检测题(一)参考答案

一、填空题

$$1. \phi(x_1, x_2, \dots, x_n) = \begin{cases} \lambda^n \exp\left\{-\lambda \sum_{i=1}^n x_i\right\}, & x_i > 0, \\ 0, & x_i = 0, \end{cases} \quad \bar{x} = 5, S^2 = 2.2$$

$$2. 0.58 \quad 3. Y \sim N(-10\mu, 250a^2)$$

$$4. \lambda = 7.68 \quad 5. U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, T = \frac{\bar{X} - \mu}{S_n/\sqrt{n}}, \chi^2 = \frac{(n-1)}{\sigma^2} S_n^2$$

二、选择题

1. A 2. D 3. B 4. D 5. A

三、解答题

$$1. F_n(x) = \begin{cases} 0, & x < 2, \\ 0.20, & 2 \leq x < 3, \\ 0.40, & 3 \leq x < 5, \\ 0.60, & 5 \leq x < 7, \\ 0.80, & 7 \leq x < 8, \\ 1, & x \geq 8. \end{cases}$$

$$2. E\bar{X} = \frac{1}{\lambda},$$

$$ES^2 = \left(1 - \frac{1}{n}\right) \frac{1}{\lambda^2}.$$

3. 对于任意的 $x_1, x_2, \dots, x_n$, 有

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \dots f(x_n; \theta) = \begin{cases} (\theta+1)^n (x_1 \dots x_n)^\theta, & \text{若 } 0 \leq x_1, \dots, x_n \leq 1, \\ 0, & \text{其他.} \end{cases}$$

$$4. p\{k > \lambda\} = 1 - p\{k \leq \lambda\} = 1 - 0.05 = 0.95.$$

$$5. n = 27.$$

$$6. Y = \frac{(X_1 + X_2)^2}{(X_1 - X_2)^2} \sim F(1, 1).$$

检测题(二)

一、填空题

1. 设随机变量 X 和 Y 相互独立且都服从正态分布 $N(0, \beta^2)$, 而 $X_1, X_2, \dots, X_9$ 和 $Y_1, Y_2, \dots, Y_9$ 是分别来自总体 X 和 Y 的简单随机变量样本, 则统计量

$$U = \frac{X_1 + X_2 + \dots + X_n}{\sqrt{Y_1^2 + Y_2^2 + \dots + Y_n^2}}$$

服从_____分布,自由度为_____.

2. 设 $X \sim N(\mu, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 且 X 、 Y 相互独立, 样本方差分别为 S_1^2, S_2^2 , 则在_____条件下, 常用的统计量 $F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$.

3. 设 $X \sim N(1, 4)$, $Y \sim N(0, 16)$, $X \sim N(4, 9)$, X, Y 相互独立, 设 $U = 4X + 3Y - X$, 则 $E(2\mu - 3) = \underline{\hspace{2cm}}$, $D(4U - 7) = \underline{\hspace{2cm}}$.

4. 设 X_1, X_2, X_3, X_4 是来自正态分布 $N(0, \sigma^2)$ 的样本, 则

$$U = \frac{\sqrt{3}X_1}{\sqrt{X_2^2 + X_3^2 + X_4^2}} \text{ 服从 } \underline{\hspace{2cm}} \text{ 分布, } EX = \underline{\hspace{2cm}}, DU = \underline{\hspace{2cm}}.$$

5. 设随机变量 X 服从分布 $t(10)$, 已知 $P\{X^2 \leq \lambda\} = 0.05$, 则 $\lambda = \underline{\hspace{2cm}}$.

二、解答题

1. 样本 $X_1, \dots, X_n (n > 1)$ 来自总体 X , 试求常数 c , 使统计量 $\frac{c(X_1 + X_2)}{\sqrt{X_3^2 + X_4^2 + X_5^2}}$ 服从 t 分布.

2. 设 $X_1, X_2, \dots, X_n$ 是参数为 λ 的泊松分布总体的一个样本, $\bar{X}, S^2$ 为样本的均值与方差, 求 $D(\bar{X}), E(S^2)$.

3. 设总体 X 服从分布 $F(n, m)$. 求证:

$$P\{X \leq 1\} = P\{X \geq 1\} = 0.5.$$

4. 从正态总体 $N(3.4, \sigma^2)$ 中抽取容量为 n 的样本, 如果要求其样本均值位于 $(1.4, 5.4)$ 内的概率不小于 0.95. 问样本容量 n 至少应取多大?

5. 设 (X_1, X_2) 来自总体 X 且服从柯西分布, 概率密度为 $f(x) = \frac{1}{\pi(4x^2)}, -\infty < x < +\infty$, 求样本均值 $\bar{X} = \frac{X_1 + X_2}{2}$ 的分布.

6. $\bar{X}$ 和 $\bar{Y}$ 是来自正态总体 $N(\mu, \sigma^2)$ 的容量为 n 的两个样本的均值. 试确定 n 使两个样本均值之差超过 σ 的概率大约为 0.01.

检测题(二)参考答案

一、填空题

$$1. t, 9 \quad 2. \sigma_1^2 = \sigma_2^2, F = \frac{S_1}{S_2} \quad 3. -3, 3, 472 \quad 4. t, 0, 3 \quad 5. \lambda = 0.004$$

二、解答题

$$1. c = \sqrt{\frac{3}{2}}. \quad 2. \frac{\lambda}{n}, \lambda. \quad 4. n \text{ 至少应取 } 35. \quad 5. f_X(x) = \frac{1}{\pi(1+x^2)}.$$

$$6. n < 14.$$

第 6 章 参数估计

6.1 参数的点估计

6.1.1 知识要点

参数的点估计及其求法.

设总体 X 的分布中含未知(待估)参数 θ , $(X_1, X_2, \dots, X_n)$ 是来自总体 X 的随机样本. 所谓点估计, 就是引入一个适当的统计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 作为待估参数 θ 的估计. 于是, 用来估计 θ 的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 称为它的估计量. 而对应于一组样本值 $(x_1, x_2, \dots, x_n)$ $\hat{\theta}(x_1, x_2, \dots, x_n)$ 称为 θ 的估计值.

特别需要指出的是, 估计量是不含未知参数的样本函数, 它是一个随机变量, 而估计值却是一个常量. 在不发生误会的情况下, 常把它们统称为估计.

用一个统计量估计未知参数的方法, 称为点估计法. 其中常用的是矩估计法和极大似然估计法两种.

1. 矩估计法只用样本矩作为相应(同类同阶)总体矩的估计

在矩估计法中, 常用的法则与前一阶矩有关的两项, 即

$\hat{u} = \bar{X}$ ——用样本均值 $\bar{X}$ 作为总体均值 u 的估计;

$\hat{\sigma}^2 = S^2$ ——用样本方差 S^2 作为总体方差 σ^2 的估计.

矩估计法的理论依据是大数定律, 它有思路清晰, 简便易行的优点. 因而这一古老方法沿用至今.

2. 极大似然估计法是基于极大似然原理的估计方法

极大似然原理的直观想法是把试验中发生概率最大的事件推断为最有可能发生.

依据这一原理引入的极大似然估计法的主要步骤如下:

①基于总体分布, 在样本点 $(x_1, x_2, \dots, x_n)$ 上建立体现概率的似然函数 $L(x_1, x_2, \dots, x_n; \theta)$, 即为样本的联合分布;

②考虑到 $\ln L$ 与 L 同时达到极大, 故为计算方便, 引入对数似然函数 $\ln L(x_1, x_2, \dots, x_n; \theta)$;

③通常, 从对数似然函数出发, 对待估参数(向量) θ 求(偏导)得似然方程(组) $\partial \ln L / \partial \theta = 0$, 并求解;

④将解中的诸 x_i 改写成 X_i , 便可得到极大似然估计量.

极大似然估计法的总体分布已知为前提. 因而, 由此为待估参数提供的信息, 一般来讲要

多于以数字特征为前提的矩估计量,所以在统计实践中广为采用.相对于矩估计,它有方法较为繁琐的不足.

此外,若 $f(\theta)$ 是 θ 的函数且具有单值反函数, $\hat{\theta}$ 又是 θ 的极大似然估计,则 $f(\hat{\theta})$ 一定也是 $f(\theta)$ 的极大似然估计.

6.1.2 典型题型解析

例 6.1.1 设总体 X 有分布列

Z	1	2	3
p	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中 $0 < \theta < 1$ 为待估参数,又设 $(X_1, X_2, \dots, X_n)$ 为总体 X 的样本,试求 θ 的矩估计量,并就样本值 (3, 1, 2, 2, 3, 2) 求 θ 的估计值.

解 为求矩估计量 $\hat{\theta}$, 要从求 EX 开始, 即

$$u = EX = \theta^2 + 4\theta(1-\theta) + 3(1-\theta)^2 = 3 - 2\theta.$$

由矩估计法知 $\hat{u} = \bar{X}$, 于是 $3 - 2\hat{\theta} = \bar{X}$, 从而解得

$$\hat{\theta} = (3 - \bar{X})/2,$$

对于给定样本值算得 $\bar{x} = 2.1677$, 故 θ 的估计值为

$$\hat{\theta} = (3 - \bar{x})/2 = 0.4167.$$

注 矩估计法的两个常用的法则是通过总体均值、总体方差来表达的. 因而当总体分布给定时, 首先要将待估参数置身于总体均值 u 或总体方差 σ^2 中, 然后按法则求解.

需要指出的是, 只有一个待估参数时, 一般只由求总体均值入手. 如果有两个待估参数时, 则需要同时求总体方差,

例 6.1.2 设总体 X 有分布密度

$$p(x) = \frac{1}{2\theta} e^{-|x|/\theta}, \quad -\infty < x < \infty,$$

其中 $\theta > 0$ 为待估参数, 试求 θ 的矩估计量.

解 因为 $u = EX = \int_{-\infty}^{+\infty} x \frac{1}{2\theta} e^{-|x|/\theta} dx = 0$. 故无法运用 $\hat{u} = \bar{X}$ 来求 $\hat{\theta}$. 为此就从方差入手, 即

$$\begin{aligned} \sigma^2 = DX = EX^2 &= \int_{-\infty}^{+\infty} x^2 \frac{1}{2\theta} e^{-|x|/\theta} dx = \frac{1}{\theta} \int_0^{+\infty} x^2 e^{-x/\theta} dx = \theta^2 \int_0^{+\infty} \left(\frac{x}{\theta}\right)^2 e^{-x/\theta} d(x/\theta) \\ &= \theta^2 \Gamma(3) = 2\theta^2, \end{aligned}$$

由矩估计法知 $\hat{\sigma}^2 = S^2$, 于是 $2\hat{\theta}^2 = S^2$, 从而解得

$$\hat{\theta} = \sqrt{S^2/2}.$$

注 这是一道从方差入手求矩估计的题目, 属非常见类型, 而求方差借助 Γ 函数进行积分计算, 是完成这类积分的常用工具.

例 6.1.3 设总体 X 有分布密度

$$p(x) = \begin{cases} 2x/\theta^2, & 0 < x < \theta, \\ 0, & \text{其他,} \end{cases}$$

其中 θ 为待估参数, $(X_1, X_2, \dots, X_n)$ 为样本, 试求 θ 的矩估计量.

解 为求矩估计量 $\hat{\theta}$, 先求 EX , 即

$$u = E(X) = \int_0^{\theta} x \frac{2x}{\theta^2} dx = \frac{2}{3} \theta.$$

由矩估计法知 $\hat{u} = \bar{X}$, 于是 $2\hat{\theta}/3 = \bar{X}$, 从而解得

$$\hat{\theta} = 3\bar{X}/2.$$

例 6.1.4 设总体 X 在 $[a, b]$ 上服从均匀分布, 其中 $a, b (b > a > 0)$ 为待估参数, 试求 a, b 的矩估计.

解 均匀分布是常用分布, 知 $E(X) = \frac{a+b}{2}$, $D(X) = (b-a)^2/12$.

由矩估计法和 $\hat{u} = \bar{X}$, $\hat{\sigma}^2 = S^2$. 于是

$$(\hat{a} + \hat{b})/2 = \bar{X}, (\hat{b} - \hat{a})^2/12 = S^2,$$

从而解得 $\hat{a} = \bar{X} - \sqrt{3}S$, $\hat{b} = \bar{X} + \sqrt{3}S$.

例 6.1.5 从某车间某天生产的滚珠中随机抽出 6 个, 测得其直径(单位: mm)为

14.6 15.1 14.9 14.8 15.2 15.1

试估计该天生产滚珠直径的均值 u 和方差 σ^2 .

解 本题仍按矩估计法处理, 解得

$$\hat{u} = \bar{x} = \frac{1}{6} \sum_{i=1}^6 x_i = 14.95,$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{6} \sum_{i=1}^6 (x_i - \bar{x})^2 = 0.0425.$$

注 运用估计式 $\hat{u} = \bar{x}$, $\hat{\sigma}^2 = S^2$, 求总体参数在指定样本点上的估计值, 是矩估计法的重要应用, 注意到不同情况下处理途径是有差别的:

- ① 如果总体分布一无所知, 那么被估参数只能是总体均值和总体方差, 本例即属此类型;
- ② 如果被估参数并非是总体均值和总体方差, 通常情况下必须以总体分布已知为前提.

例 6.1.6 设总体 X 的分布律为

$$P\{X=m\} = q^{m-1}p, m=1, 2, 3, \dots,$$

其中 $0 < q < 1 (p=1-q)$ 为待估参数, 试求 q 的矩估计和极大似然估计.

解 为求矩估计 $\hat{q}$, 求 EX , 即

$$\begin{aligned} u = EX &= \sum_{m=1}^{\infty} m q^{m-1} p = p \sum_{m=1}^{\infty} (q^m)' = p \left(\sum_{m=1}^{\infty} q^m \right)' = p [q/(1-q)] \\ &= p/(1-q)^2 = 1/(1-q). \end{aligned}$$

由矩估计法知 $\hat{u} = \bar{X}$, 于是 $1/(1-\hat{q}) = \bar{X}$, 从而解得

$$\hat{q} = 1 - 1/\bar{X}$$

为求极大似然估计 $\hat{q}$, 从列出样本点 $(x_1, x_2, \dots, x_n)$ 上的似然函数开始, 即

$$L(x_1, x_2, \dots, x_n; q) = \prod_{i=1}^n q^{x_i-1} p = p^n q^{\sum_{i=1}^n (x_i-1)} = (1-q)^n q^{\sum_{i=1}^n x_i - n},$$

两边取对数得对数似然函数

$$\ln L = n \ln(1-q) + \left[\left(\sum_{i=1}^n x_i \right) - n \right] \ln q.$$

求导得似然方程并求解

$$\frac{d\ln L}{dq} = -\frac{n}{1-q} + \frac{(\sum_{i=1}^n x_i) - n}{q}.$$

$$\text{令 } \frac{d\ln L}{dq} = 0 \text{ 则 } \frac{q}{1-q} = \frac{(\sum_{i=1}^n x_i) - n}{n}, q = 1 - \frac{1}{\sum_{i=1}^n x_i / n}.$$

把上述表达式中的 x_i 改写 X_i , 得 q 的极大似然估计

$$\hat{q} = 1 - \frac{1}{\sum_{i=1}^n X_i / n} = 1 - 1/\bar{X},$$

其结果与矩估计相同, 即 $q = 1 - 1/\bar{X}$.

例 6.1.7 已知总体 X 的分布密度为

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

其中 $\theta > 0$ 为待估参数, 试求 θ 的矩估计和极大似然估计.

解 先求矩估计 $\hat{\theta}$.

$$\begin{aligned} u = EX &= \int_0^{+\infty} x \theta^2 x e^{-\theta x} dx \\ &= \frac{1}{\theta} \int_0^{+\infty} (\theta x)^2 e^{-\theta x} d(\theta x) \\ &= \frac{1}{\theta} \Gamma(3) = \frac{2}{\theta}. \end{aligned}$$

由矩估计法知 $\hat{u} = \bar{X}$, 于是 $\frac{2}{\hat{\theta}} = \bar{X}$, 从而解得

$$\hat{\theta} = 2/\bar{X}.$$

下面继续求极大似然估计 $\hat{\theta}$.

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n \theta^2 x_i e^{-\theta x_i} = \theta^{2n} \left(\prod_{i=1}^n x_i \right) e^{-\theta \sum_{i=1}^n x_i},$$

$$\ln L = 2n \ln \theta + \ln \prod_{i=1}^n x_i - \theta \sum_{i=1}^n x_i,$$

$$\frac{d\ln L}{d\theta} = \frac{2n}{\theta} - \sum_{i=1}^n x_i.$$

$$\text{令 } \frac{d\ln L}{d\theta} = 0 \text{ 则 } \frac{2n}{\theta} = \sum_{i=1}^n x_i \Rightarrow \theta = 2n / \sum_{i=1}^n x_i.$$

把上述表达式中的 x_i 改写成 X_i , 得 θ 的极大似然估计为

$$\hat{\theta} = 2n / \sum_{i=1}^n X_i = 2/\bar{X}.$$

注 回顾两种点估计, 对其方法要点及其利弊做一简要综合.

(1) 矩估计法是从数字特征出发, 操作方法方便简捷, 数学原理简明扼要, 但具体计算中可能会遇到某些技巧, 如前例中级数求和的逐项微分和利用 Γ 函数.

(2)极大似然估计法以总体分布已知为前提,虽然方法甚为繁琐,数学原理也不如矩估计法简明直观,但由于似然函数集中了总体分布较充分的信息,因而得到的估计量较矩估计量有较多的优良性.它是目前应用非常广泛的一种估计方法.

此外,本题演算结果两种估计方法是相等的.需要指出的是,矩估计与极大似然估计的相等纯属巧合,不相等的情形也相当普遍.

例 6.1.8 已知总体 X 的分布密度为

$$p(x) = \begin{cases} \frac{x}{2\theta^2} e^{-x^2/\theta^2}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

其中 $\theta > 0$ 为待估参数,试求 θ 的矩估计和极大似然估计.

解 先求矩估计

$$u = EX = \int_0^{+\infty} x \frac{x}{2\theta^2} e^{-x^2/\theta^2} dx = \frac{1}{2\theta} \cdot \frac{\theta^2}{2} \int_0^{+\infty} \left(\frac{x^2}{\theta^2}\right)^{\frac{1}{2}} e^{-x^2/\theta^2} d\left(\frac{x^2}{\theta^2}\right) = \frac{\theta}{4} \Gamma\left(\frac{3}{2}\right) = \frac{\theta\sqrt{\pi}}{8}.$$

由矩估计法知 $\hat{u} = \bar{X}$, 于是 $\frac{1}{8}\hat{\theta}\sqrt{\pi} = \bar{X}$, 从而解得

$$\hat{\theta} = 8\bar{X}/\sqrt{\pi}.$$

下面继续求极大似然估计.

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n \frac{x_i}{2\theta^2} e^{-x_i^2/\theta^2} = 2^{-n} \theta^{-2n} \left(\prod_{i=1}^n x_i \right) e^{-\frac{1}{\theta^2} \sum_{i=1}^n x_i^2}$$

$$\ln L = -n \ln 2 - 2n \ln \theta + \ln \prod_{i=1}^n x_i - \frac{1}{\theta^2} \sum_{i=1}^n x_i^2,$$

$$\frac{d \ln L}{d\theta} = \frac{-2n}{\theta} + \frac{2}{\theta^3} \sum_{i=1}^n x_i^2,$$

$$\text{令 } \frac{d \ln L}{d\theta} = 0 \text{ 则 } \frac{2n}{\theta} = \frac{2}{\theta^3} \sum_{i=1}^n x_i^2 \Rightarrow \theta = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}.$$

把上述表达式中的 x_i 改写 X_i , 得 θ 的极大似然估计为

$$\hat{\theta} = \sqrt{\sum_{i=1}^n X_i^2 / n}.$$

注 对于极大似然估计法,要掌握适当.即在引入对数似然函数后,由似然事件的情形事实上并非所有题目均可按常规方法处理.

例 6.1.9 设总体 $X \sim U[a, b]$ ($b > a > 0$), 求 a, b 的极大似然估计.

解 按常规方法写出 X 分布密度

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

之后,便有似然函数

$$L(x_1, x_2, \dots, x_n; a, b) = \frac{1}{(b-a)^n} \quad a \leq x_i \leq b, \quad i = 1, 2, \dots, n,$$

于是有似然方程组

$$\frac{\partial \ln L}{\partial a} = \frac{n}{b-a}, \quad \frac{\partial \ln L}{\partial b} = \frac{-n}{b-a}.$$

若令 $\frac{\partial \ln L}{\partial a} = 0$ 及 $\frac{\partial \ln L}{\partial b} = 0$ 均无法解出 a, b 只得回到似然函数去处理.

按极大似然原理的本意, 应在 $L(x_1, x_2, \dots, x_n; a, b) = 1/(b-a)^n$ 达到极大下定出 a, b . 即相当于要求 $(b-a)^n$ 取极小. 按通常理解, 此时 b 应尽量小, a 应尽可能大. 它们又共同受到 $a \leq x_i \leq b$ 的约束. 而 b 应尽量小又不能小于任一个 x_i , 故 b 只能是 x_i 中最大的一个. 即 $\hat{b} = \max\{x_1, x_2, \dots, x_n\}$. 同理 a 应尽量大又不能大于任一个 x_i , 故 a 只能是 x_i 中最小的一个. 即 $\hat{a} = \min\{x_1, x_2, \dots, x_n\}$. 从某种意义上讲, 这一结果与矩估计相比更符合实际.

例 6.1.10 设总体 X 有分布列.

$$\begin{pmatrix} 1 & 2 & 3 \\ \theta^2 & 2\theta(1-\theta) & (1-\theta)^2 \end{pmatrix}$$

其中 $0 < \theta < 1$ 为待估参数, 又设 $(X_1, X_2, \dots, X_n)$ 为总体 X 的样本. 求 θ 的极大似然估计并就样本值 $(3, 1, 2, 2, 3, 2)$ 求 θ 的估计值.

解 设在总体 X 中, 抽取容量为 n 的样本 $(X_1, X_2, \dots, X_n)$, 并设其中 n_1 个 X_i 取 1, n_2 个 X_i 取 2, 另有 $(n - n_1 - n_2)$ 个 X_i 取 3, 于是与 θ 有关的似然函数为

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \theta) &= (\theta^2)^{n_1} [2\theta(1-\theta)]^{n_2} [(1-\theta)^2]^{n-n_1-n_2} \\ &= \theta^{2n_1} 2^{n_2} \theta^{n_2} (1-\theta)^{n_2} (1-\theta)^{2n-2n_1-2n_2} \\ &= 2^{n_2} \theta^{2n_1+n_2} (1-\theta)^{2n-2n_1-n_2}, \end{aligned}$$

$$\begin{aligned} \frac{dL}{d\theta} &= 2^{n_2} (2n_1 + n_2) \theta^{2n_1+n_2-1} (1-\theta)^{2n-2n_1-n_2} - \\ &\quad 2^{n_2} \theta^{2n_1+n_2} (2n-2n_1+n_2) (1-\theta)^{2n-2n_1-n_2-1}. \end{aligned}$$

令 $\frac{dL}{d\theta} = 0$, 则 $(2n_1 + n_2)(1-\theta) = \theta(2n-2n_1-n_2) \Rightarrow \hat{\theta} = (2n_1 + n_2)/2n$.

由于题中 $n_1 = 1, n_2 = 3, n = 6$. 故 θ 的估计值为

$$\hat{\theta} = (2 \times 1 + 3)/12 = 0.4167.$$

同步练习 6.1

1. 设总体 X 有分布列

$$\begin{pmatrix} -1 & 0 & 2 \\ 2\theta & \theta & 1-3\theta \end{pmatrix}$$

其中 $0 < \theta < 1/3$ 为待估参数, 试求 θ 的矩估计量.

2. 设总体 X 有分布密度

$$p(x) = 2(\theta - x)/\theta^2, \quad 0 < x < \theta.$$

其中 $\theta > 0$ 为待估参数, 试求 θ 的矩估计.

3. 已知总体 X 有分布密度

$$p(x) = \frac{\beta^K}{(K-1)!} x^{K-1} e^{-\beta x}, \quad x > 0,$$

其中 K 为常数, β 为待估参数, 试求 β 的矩估计.

4. 已知总体 X 在 $[a-b, 3a+b]$ 上服从均匀分布, 其中 $a > 0, b > 0$ 为待估参数. 试求 $a,$

b 的矩估计.

5. 下面是一批灯泡寿命(单位: h)的测试样本:

1 458 1 395 1 562 1 614 1 490

1 382 1 536 1 496 1 351 1 478

试用矩估计法对这批灯泡的平均寿命 M 及寿命方差 σ^2 做出估计.

6. 某元件正常工作时间(单位: h)服从以 λ 为参数的指数分布. 从中抽测 50 只元件的寿命经分组后的频数分布为

寿命组中值(x_i): 550 650 750 850 950

频数: 6 11 18 13 2

试用矩估计法估计参数 λ .

7. 已知总体 $X \sim B(1, p)$, $0 < p < 1$ 为待估参数. 试求 p 的矩估计和极大似然估计.

8. 已知总体 X 有分布密度

$$p(x) = (a+1)x^a, 0 < x < 1,$$

其中 $a > 0$ 为待估参数, 试求 a 的矩估计量和极大似然估计量, 并就样本值(1.50 0.74 0.62, 0.38 0.90)求相应的估计值.

9. 设总体 $X \sim N(0, \sigma^2)$, σ^2 为待估参数, 试求 σ^2 极大似然估计.

同步练习 6.1 参考答案

1. $\hat{\theta} = (2 - \bar{X})/8.$

2. $\hat{\theta} = 3\bar{X}.$

3. $\hat{\beta} = K/\bar{X}$ (用 Γ 函数进行积分计算).

4. $\hat{a} = \bar{X}/2, \hat{b} = \sqrt{3}S - \bar{X}/2.$

5. $\hat{\mu} = \bar{x} = 1\,476.2, \hat{\sigma}^2 = S^2 = (78.73\,093)^2 = 6\,198.56.$

6. $\hat{\lambda} = 1/\bar{x} = 1/738 = 1.355 \times 10^{-3}.$

7. $\hat{p} = \bar{X}.$

8. 矩估计 $\hat{\alpha} = (2\bar{X} - 1)/(1 - \bar{X})$, 其估计值为 $\hat{\alpha} = 3.8\,140.$

似然估计 $\hat{\alpha} = -1 - n/\sum_{i=1}^n \ln X_i$, 其估计值为 $\hat{\alpha} = 2.4\,564.$

9. $\hat{\sigma}^2 = \sum_{i=1}^n X_i^2/n.$

6.2 估计量的评价标准

6.2.1 知识要点

对于估计量的优良性评判可以从不同的角度去考虑, 无偏性、有效性准则是概率意义明确、应用较为广泛的两条.

① 设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 作为待估参数 θ 的估计量. 如果 $E\hat{\theta} = \theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估

计量.

无偏性的概率意义是估计量 $\hat{\theta}$ 作为样本函数,它在历次试验中的观测值总是围绕在 θ 其值的两侧摆动.此时 $\hat{\theta}$ 观测值的平均值与待估参数 θ 的真值几乎没有差别.

② 设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是 θ 的无偏估计量.如果 $D\hat{\theta}_1 < D\hat{\theta}_2$, 则称 $\hat{\theta}_1$ 作为 θ 的估计量较 $\hat{\theta}_2$ 有效, 或说成 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 为 θ 的有效估计量.

$\hat{\theta}_1, \hat{\theta}_2$ 作为无偏估计,它们在历次试验中的观测值总是围绕 θ 真值的两侧摆动,而 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效,指的是 $\hat{\theta}_1$ 在 θ 两侧摆动的幅度比 $\hat{\theta}_2$ 更小些,即 $\hat{\theta}_1$ 的观测值较 $\hat{\theta}_2$ 更集中在 θ 的真值附近.

6.2.2 典型题型解析

例 6.2.1 设 X 的均值 $\mu = EX$ ($X_1, X_2, \dots, X_n$) 为样本.试证

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad \text{为方差 } \sigma^2 = DX \text{ 的无偏估计.}$$

证 考虑到样本与总体独立同分布的特征,故有

$$\begin{aligned} E\hat{\sigma}^2 &= E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right) = \frac{1}{n} \left[\sum_{i=1}^n E(X_i - EX_i)^2 \right] \\ &= \frac{1}{n} \sum_{i=1}^n DX_i = \sigma^2. \end{aligned}$$

注 明确样本与总体独立同分布的观念,是本题获证的关键,其证明的思路和技巧有一定的代表性,涉及 σ^2 的估计,归纳汇总如下:

在 μ 已知的条件下, σ^2 的无偏估计是 $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$,这一结果在统计推断的有关课题中多次应用.

在 μ 未知的条件下, $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 作为 σ^2 的估计是有偏的,而 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 才是 σ^2 的无偏估计.

此外,顺便指出 S^2 虽是 σ^2 的无偏估计, S 却不是 σ 的无偏估计.

例 6.2.2 设 $(X_1, X_2, \dots, X_n)$ 是来自参数为 $\lambda > 0$ 的泊松分布的样本.对于任意的 $a (0 \leq a \leq 1)$ 以及样本均值 $\bar{X}$, $S^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, 试证 $a\bar{X} + (1-a)\frac{n}{n-1}S^{*2}$ 是 λ 的无偏估计.

证 $E\bar{X} = EX = \lambda$, $ES^{*2} = \frac{n-1}{n}\sigma^2 = \frac{n-1}{n}\lambda$, 故有

$$\begin{aligned} E\left[a\bar{X} + (1-a)\frac{n}{n-1}S^{*2}\right] &= aE\bar{X} + (1-a)\frac{n}{n-1}ES^{*2} \\ &= a\lambda + (1-a)\frac{n}{n-1} \cdot \frac{n-1}{n}\lambda = \lambda. \end{aligned}$$

例 6.2.3 设总体 X 有分布律

$$P\{X=m\} = \frac{\lambda^m}{m!} e^{-\lambda}, \quad m=0, 1, 2, \dots,$$

其中 $\lambda > 0$ 为参数.试证 $\lambda(X) = (-1)^m$ 是 $e^{-2\lambda}$ 的无偏估计.

证 $\lambda(X)$ 作为随机变量 X 的函数,其数学期望为

$$\begin{aligned}
 E\lambda(X) &= \sum_{m=0}^{\infty} \lambda(X) P\{X=m\} = \sum_{m=0}^{\infty} (-1)^m \frac{\lambda^m}{m!} e^{-\lambda} \\
 &= e^{-\lambda} \sum_{m=0}^{\infty} (-1)^m \frac{\lambda^m}{m!} = e^{-\lambda} e^{-\lambda} = e^{-2\lambda}.
 \end{aligned}$$

同步练习 6.2

1. 已知总体 X 有分布密度

$$p(x) = \frac{1}{2\theta} e^{-|x|/\theta}, -\infty < x < \infty,$$

其中 $\theta > 0$ 为待估参数. 试求 θ 的极大似然估计, 并问该估计是不是 θ 的无偏估计, 为什么?

2. 设总体 $X \sim N(\mu, \sigma^2)$, 其中 μ 为待估参数, σ^2 为已知参数, (X_1, X_2, X_3) 为样本. 试考察 μ 的下列估计量的无偏性与有效性:

$$\hat{\mu}_1 = \frac{4}{3}X_1 - \frac{2}{3}X_2 + \frac{1}{3}X_3;$$

$$\hat{\mu}_2 = \frac{1}{3}X_1 + \frac{2}{3}X_2 - \frac{2}{3}X_3;$$

$$\hat{\mu}_3 = \frac{2}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3.$$

同步练习 6.2 参考答案

1. $\hat{\theta} = \sum_{i=1}^n |X_i|/n$, $\hat{\theta}$ 是 θ 的无偏估计,

$$\begin{aligned}
 E\hat{\theta} &= \frac{1}{n} \sum_{i=1}^n E|X_i| = \frac{1}{n} \sum_{i=1}^n E|X| = E|X| \\
 &= \frac{1}{2\theta} \int_{-\infty}^{+\infty} |x| e^{-|x|/\theta} dx = \frac{1}{\theta} \int_0^{+\infty} x e^{-x/\theta} dx \\
 &= \theta \int_0^{+\infty} \left(\frac{x}{\theta}\right) e^{-x/\theta} d\left(\frac{x}{\theta}\right) = \theta \Gamma(2) = \theta.
 \end{aligned}$$

2. $\hat{\mu}_1, \hat{\mu}_3$ 都是 μ 的无偏估计量, $\hat{\mu}_3$ 较 $\hat{\mu}_1$ 为 μ 的有效估计量.

6.3 区间估计

6.3.1 知识要点

1. 置信区间与置信度

定义 设总体的 X 分布函数 $F(x; \theta)$ 含有一未知参数 θ . 对于给定值 $\alpha (0 < \alpha < 1)$, 若由样本 $X_1, X_2, \dots, X_n$ 确定的两个统计量 $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$ 和 $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$ 满足

$$P\{\underline{\theta}(X_1, X_2, \dots, X_n) < \theta < \bar{\theta}(X_1, X_2, \dots, X_n)\} = 1 - \alpha,$$

则称随机区间 $(\underline{\theta}, \bar{\theta})$ 是 θ 的置信度为 $1 - \alpha$ 的置信区间, $\underline{\theta}$ 和 $\bar{\theta}$ 分别称为置信度为 $1 - \alpha$ 的双

侧置信区间的置信下限和置信上限 $1 - \alpha$ 称为置信度.

2. 一个正态总体参数的区间估计

假设总体 $X \sim N(\mu, \sigma^2)$, $(X_1, X_2, \dots, X_n)$ 是来自总体 X 的简单随机样本, $\bar{X}$ 是样本均值, S^2 是样本方差. 参数的区间估计如下表:

	被估参数	条件	用于估计的随机变量(样本函数)	随机变量分布	置信度为 $1 - \alpha$ 的置信区间
单个正态总体	总体均值 μ	σ^2 已知	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$N(0, 1)$	$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \right)$
		σ^2 未知	$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$	$t(n-1)$	$\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \right)$
	总体方差 σ^2	μ 已知	$\chi^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$	$\chi^2(n)$	$\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)} \right)$
		μ 未知	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ $= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$	$\chi^2(n-1)$	$\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \right)$ 或 $\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \right)$

3. 两个正态总体参数的区间估计

假设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$; $(X_1, X_2, \dots, X_{n_1})$ 和 $(Y_1, Y_2, \dots, Y_{n_2})$ 分别是来自总体 X 和 Y 的简单随机样本, $\bar{X}, S_1^2, \bar{Y}, S_2^2$ 是相应的样本均值和样本方差. 参数的区间估计如下表:

	被估参数	条件	用于估计的随机变量及分布	置信度为 $1 - \alpha$ 的置信区间
两个正态总体	$\mu_1 - \mu_2$	两样本独立, 且 σ_1^2, σ_2^2 已知	$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	$\left(\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}, \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \right)$
		两样本独立, 且 σ_1^2, σ_2^2 未知, 但知相等	$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$ $S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	$\left(\bar{X} - \bar{Y} - t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$
		σ_1, σ_2 未知, 但 n_1, n_2 很大 (> 5)	$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim N(0, 1)$	$\left(\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$
	$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1, μ_2 未知	$F = \frac{\frac{(n_1 - 1)S_1^2}{\sigma_1^2} / (n_1 - 1)}{\frac{(n_2 - 1)S_2^2}{\sigma_2^2} / (n_2 - 1)} \sim F(n_1 - 1, n_2 - 1)$	$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)}, \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)} \right)$

6.3.2 典型题型解析

例 6.3.1 从一大批同型号金属线中,随机选取 10 根,测得它们的直径(单位: mm)为

1.23 1.24 1.26 1.29 1.20 1.32 1.23 1.23 1.29 1.28

(1)如果金属线直径 $x \sim N(\mu, 0.04^2)$,试求平均直径 μ 在置信度 $1 - \alpha = 0.95$ 下的置信区间;

(2)另设金属线直径 $x \sim N(\mu, \sigma^2)$,其中 σ 为未知参数,试求平均直径 μ 在置信度 $1 - \alpha = 0.95$ 下的置信区间.

解 本题为正态总体均值的区间估计的两种类型.为了后面的需要先计算均值与方差,得 $\bar{x} = 1.257, s^2 = 0.0014$.

(1)因 $\sigma^2 = 0.04^2$ 已知,故选 $U = (\bar{x} - \mu) / \sqrt{\sigma^2 / n} \sim N(0, 1)$ 为估计用样本函数.由 $\alpha = 0.05$ 查正态分布表得临界值

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96.$$

算得置信下限、置信上限分别为

$$\underline{\theta} = \bar{x} - z_{\frac{\alpha}{2}} \sqrt{\sigma^2 / n} = 1.257 - 1.96 \sqrt{0.04^2 / 10} = 1.257 - 0.0248 = 1.2322,$$

$$\bar{\theta} = \bar{x} + z_{\frac{\alpha}{2}} \sqrt{\sigma^2 / n} = 1.257 + 0.0248 = 1.2818.$$

由此得 μ 的置信区间(1.2322, 1.2818).

(2)题设中 σ 未知,故选 $t = (\bar{X} - \mu) / \sqrt{s^2 / n} \sim t(n - 1)$ 为估计用样本函数.

由 $\alpha = 0.05$ 查 t 分布表得临界值

$$t_{\frac{\alpha}{2}}(n - 1) = t_{0.025}(9) = 2.2622.$$

算得置信下限、置信上限分别为

$$\underline{\theta} = \bar{x} - t_{\frac{\alpha}{2}} \sqrt{s^2 / n} = 1.257 - 2.2622 \sqrt{0.0014 / 10} = 1.257 - 0.0268 = 1.2302,$$

$$\bar{\theta} = \bar{x} + t_{\frac{\alpha}{2}} \sqrt{s^2 / n} = 1.257 + 0.0268 = 1.2838.$$

由此便得置信度 $1 - \alpha = 0.95$ 时 μ 的置信区间(1.2302, 1.2838).

例 6.3.2 已知总体 $X \sim N(\mu, \sigma^2)$,试分别在下列条件下求指定参数的置信区间:

(1) σ^2 未知, $n = 20, \bar{x} = 13.2, s^2 = 5, \alpha = 0.05$ 求 μ 的置信区间;

(2) μ 未知, $n = 12, s^2 = 1.243, \alpha = 0.02$ 求 σ^2 的置信区间.

解 (1) σ^2 未知,故选 $t = (\bar{X} - \mu) / \sqrt{s^2 / n} \sim t(n - 1)$ 为估计用样本函数.

由 $\alpha = 0.05$ 查 t 分布表得临界值

$$t_{\frac{\alpha}{2}}(n - 1) = t_{0.025}(19) = 2.0860,$$

算得置信下限、置信上限分别为

$$\begin{aligned} \underline{\theta} &= \bar{x} - t_{\frac{\alpha}{2}} \sqrt{s^2 / n} \\ &= 13.2 - 2.0860 \sqrt{5 / 20} = 13.2 - 1.043 = 12.157, \end{aligned}$$

$$\bar{\theta} = \bar{x} + t_{\frac{\alpha}{2}} \sqrt{s^2 / n} = 13.2 + 1.043 = 14.243.$$

由此便得置信度 $1 - \alpha = 0.95$ 时 μ 的置信区间为(12.157, 14.243).

(2)由 μ 未知, 故选 $\chi^2 = (n-1)s^2/\sigma^2 \sim \chi^2(n-1)$ 为估计用样本函数. 由 $\alpha=0.02$ 查 χ^2 分布表得临界值

$$\lambda_2 = \chi_{\frac{\alpha}{2}}^2(n-1) = \chi_{0.01}^2(11) = 24.725,$$

$$\lambda_1 = \chi_{1-\frac{\alpha}{2}}^2(n-1) = \chi_{0.99}^2(11) = 3.053.$$

算得置信下限、置信上限分别为

$$\underline{\theta} = (n-1)s^2/\lambda_2 = 11 \times 1.243/24.725 = 0.553,$$

$$\bar{\theta} = (n-1)s^2/\lambda_1 = 11 \times 1.243/3.053 = 4.479.$$

由此便得置信度 $1-\alpha=0.98$ σ^2 的置信区间(0.553, 4.479).

例 6.3.3 设灯泡寿命(单位: h)服从正态分布. 今从一大批同类灯泡中随机抽取其中 10 只, 测得样本方差 $s^2=400$. 试求这批灯泡寿命方差 σ^2 的置信度为 90% 的置信区间.

解 题设中未提及总体均值是否已知, 对此通常按未知处理, 故选 $\chi^2 = (n-1)s^2/\sigma^2 \sim \chi^2(n-1)$ 为估计用样本函数. 由 $\alpha=0.1$ 查 χ^2 分布表得临界值

$$\lambda_2 = \chi_{\frac{\alpha}{2}}^2(n-1) = \chi_{0.05}^2(9) = 16.919,$$

$$\lambda_1 = \chi_{1-\frac{\alpha}{2}}^2(n-1) = \chi_{0.95}^2(9) = 3.325.$$

算得置信下限、置信上限分别为

$$\underline{\theta} = (n-1)s^2/\lambda_2 = 9 \times 400/16.919 = 212.78,$$

$$\bar{\theta} = (n-1)s^2/\lambda_1 = 9 \times 400/3.325 = 1082.71.$$

由此便得置信度为 90% σ^2 的置信区间为(212.78, 1082.71).

例 6.3.4 从一批钢索中抽样 10 根, 测得其折断力(单位: kg)为

578 572 570 568 572 570 570 596 584 572

若折断力 $X \sim N(\mu, \sigma^2)$, $\mu=580$. 试求方差 σ^2 、均方差 σ 的置信度为 0.95 的置信区间.

解 题设中 μ 已知, 故选 $\chi^2 = \sum_{i=1}^n (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$ 为估计用样本函数.

由 $\alpha=0.05$ 查 χ^2 分布表得临界值

$$\lambda_2 = \chi_{\frac{\alpha}{2}}^2(n) = \chi_{0.025}^2(10) = 20.483,$$

$$\lambda_1 = \chi_{1-\frac{\alpha}{2}}^2(n) = \chi_{0.975}^2(10) = 3.247.$$

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)^2 &= \sum_{i=1}^n x_i^2 - 2n\mu\bar{x} + n\mu^2 \\ &= 3309232 - 2 \times 10 \times 580 \times 575.2 + 10 \times 580^2 = 912. \end{aligned}$$

算得置信下限、置信上限分别为

$$\underline{\theta} = \sum_{i=1}^n (X_i - \mu)^2/\lambda_2 = 912/20.483 = 44.52,$$

$$\bar{\theta} = \sum_{i=1}^n (X_i - \mu)^2/\lambda_1 = 912/3.247 = 280.87.$$

于是得置信度 $1-\alpha=0.95$ σ^2 的置信区间(44.52, 280.87). σ 的置信区间($\sqrt{44.52}$, $\sqrt{280.87}$), 即(6.67, 16.76).

例 6.3.5 随机地从 A 组导线中抽取 4 根, 从 B 组导线中抽取 5 根, 测得其电阻(Ω)为

A 组导线 0.143 0.142 0.143 0.137

B 组导线 0.140 0.142 0.136 0.138 0.140

试测试数据分别服从正态分布 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$, 且它们相互独立, 又 μ_1, μ_2, σ^2 均未知, 试求 $\mu_1 - \mu_2$ 的 95% 的置信区间.

解 σ_1^2, σ_2^2 未知, 但知相等: $\sigma_1^2 = \sigma_2^2 = \sigma^2$. 求两正态总体期望差 $\mu_1 - \mu_2$ 的置信区间应选用随机变量

$$t = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2),$$

其中
$$S_w = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}.$$

由 $P[|t| < t_{\alpha/2}(n_1 + n_2 - 2)] = P[|t| < t_{0.025}(4 + 5 - 2)] = 0.95$, 查自由度为 7 的 t 分布表得分位数 $t_{0.025}(7) = 2.365$, 于是得 $\mu_1 - \mu_2$ 的置信度为 0.95 的置信区间为

$$\left(\bar{X} - \bar{Y} - 2.365 S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X} - \bar{Y} + 2.365 S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right).$$

易算得 $\bar{x} = \left(\sum_{i=1}^4 x_i \right) / 4 = 0.14125$, $\bar{y} = \left(\sum_{i=1}^5 y_i \right) / 5 = 0.1392$,

$$s_1^2 = 8.24 \times 10^{-6}, s_2^2 = 5.992 \times 10^{-6}, 2.365 S_w \sqrt{1/n_1 + 1/n_2} = 0.0042.$$

于是 $\mu_1 - \mu_2$ 的置信度为 0.95 的置信区间为 $(-0.00214, 0.01625)$.

同步练习 6.3

1. 从一大批零件中抽取 6 个, 测得其长度(单位: mm)为

14.70 15.21 14.90 15.32 14.91 15.32

(1) 如果零件长度 $X \sim N(\mu, 0.5)$, 试求 μ 的置信度 $1 - \alpha = 0.90$ 的置信区间;

(2) 如果零件长度 $X \sim N(\mu, \sigma^2)$, 其中 σ^2 为未知参数, 试求 μ 的置信度 $1 - \alpha = 0.90$ 的置信区间.

2. 设某种品牌的灯泡寿命(单位: h) $X \sim N(\mu, \sigma^2)$, 从中抽测 9 只, 测得平均寿命 $\bar{x} = 1000$. 寿命方差 $s^2 = 100^2$, 在信度 $\alpha = 0.05$ 下, 试求总体均值 μ , 总体方差 σ^2 及总体均方差 σ 的置信区间.

3. 已知容量 $n = 16$ 的某样本, 其样本均值 $\bar{x} = 196$. 样本修正方差 $s^2 = 54^2$. 在 $\alpha = 0.01$ 下, 试求 μ, σ^2 及 σ 的置信区间.

4. 已知 (X_1, X_2, X_3, X_4) 为正态总体 X 的一个样本, $(12.6, 13.4, 12.8, 13.2)$ 是它的一组观测值, 在置信度 $1 - \alpha = 0.95$ 以及下述指定条件下, 求 σ^2 的置信区间:

(1) 总体 $X \sim N(\mu, \sigma^2)$; (2) 总体 $X \sim N(12.5, \sigma^2)$

5. 已知总体 $X \sim N(\mu, \sigma^2)$, μ 为待估参数, $\sigma^2 = 37$. 为使 $\alpha = 0.01$, μ 的置信区间长度不超过 7.3. 试问样本容量 n 最小应是多少?

同步练习 6.3 参考答案

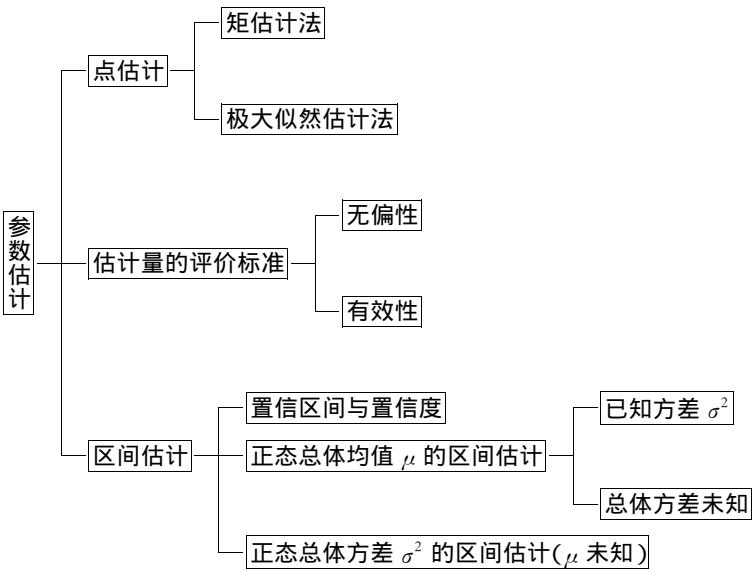
1. (1) (14.59, 15.53). (2) (14.85, 15.27).

- 2. μ (918.47 ,1 081.53) ,
 σ^2 (5 132.59 ,41 284.40) , σ (71.64 ,203.19).
- 3. μ (156.22 ,235.78) ,
 σ^2 (1 333.50 ,9 506.63) , σ (36.52 ,97.50).
- 4. (1)(0.04 ,1.85) (2)(0.13 ,2.89)
- 5. $n = [4 \times 37^2 \times 2.575^2 / 7.3^2 + 1] = [681.35 + 1] = 682$.

本章常考题型归纳

考 点	常见题型	例 题
参数的点估计	参数矩估计	例 6.1.1 例 6.1.2 例 6.1.3 例 6.1.4 例 6.1.5
	参数的极大似然估计	例 6.1.6 例 6.1.7 例 6.1.8
	非常规题型	例 6.1.9 例 6.1.10
估计量的评价标准	无偏性与有效性	例 6.2.1 例 6.2.2 例 6.2.3
区间估计	已知方差 σ^2 均值 μ 的区间估计	例 6.3.1
	未知方差 σ^2 均值 μ 的区间估计	例 6.3.1 例 6.3.2
	μ 未知 ,总体方差 σ^2 的区间估计	例 6.3.2 例 6.3.3
	μ 已知 ,总体方差 σ^2 的区间估计	例 6.3.4

本章知识结构网络



检测题(一)

1. 设总体 X 服从二项分布 $B(k, p)$, k 是正整数 $0 < p < 1$,两者都是未知参数 ,试求 k 和 p 的矩估计.

2. 设总体 X 具有分布密度 $P(x_i, a) = (a+1)x^a$, $0 < x < 1$, 其中 $a > -1$ 是未知参数, $x_1, \dots, x_n$ 为一个样本, 试求参数 a 的矩估计和极大似然估计.

3. 设总体 $X \sim N(\mu, \sigma^2)$, X_1, X_2, X_3 是一个样本, 试验证:

$$\hat{\mu}_1 = \frac{1}{5}X_1 + \frac{3}{10}X_2 + \frac{1}{2}X_3,$$

$$\hat{\mu} = \frac{1}{3}X_1 + \frac{1}{4}X_2 + \frac{5}{12}X_3,$$

$$\hat{\mu} = \frac{1}{3}X_1 + \frac{1}{6}X_2 + \frac{1}{2}X_3$$

都是 μ 的无偏估计量, 并分析哪一个最好.

4. 设总体 $X \sim B(m, p)$, 试求 p 的极大似然估计量.

5. 设总体 X 的密度函数为

$$f(x) = \begin{cases} (a+1) \cdot x^a, & 0 < x < 1, \\ 0, & \text{其他}, \end{cases} \quad a > -1.$$

求 a 的极大似然估计量和矩估计量.

6. 设 $X_1, X_2, \dots, X_n$ 是总体 $X \sim N(\mu, \sigma^2)$ 的一个样本, 已知方差 σ^2 , 试对期望做区间估计 (置信度为 $1 - \alpha$).

7. 一个车间生产滚珠, 滚珠直径服从正态分布, 从某天的产品里随机抽出 5 个量得直径如下 (单位: mm):

14.6 15.1 14.9 15.2 15.1

如果知道该天直径的方差是 0.05, 试找出平均直径的置信区间 ($\alpha = 0.05$).

8. 已知某种木材横纹抗压力的实验值服从正态分布, 对 10 个试件做横纹抗压力试验得数据如下 (单位: MPa):

48.2 49.3 45.7 47.1 51.0 44.6 43.5 41.8 39.4 46.9

试对该木材平均横纹抗压力进行区间估计 ($\alpha = 0.05$).

9. 岩石密度的测量误差服从正态分布, 随机抽测 12 个样本得 $s = 0.2$, 求 σ^2 的置信区间 ($\alpha = 10\%$).

检测题(一)参考答案

$$1. \hat{p} = \frac{\bar{X} - \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}{x},$$

$$\hat{k} = \left[\frac{\bar{X}^2}{\bar{X} - \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right].$$

$$2. \text{由 } \bar{X} = \frac{\alpha+1}{\alpha+2} \text{ 解出 } \hat{\alpha} = \frac{\bar{X}-1}{1-\bar{X}}.$$

$$\hat{\alpha} = - \left[1 + \frac{n}{\sum_{i=1}^n \ln X_i} \right].$$

3. $\hat{\mu}_2$ 最好.

$$4. \hat{p} = \frac{\bar{x}}{m}.$$

从而得到 p 的最大似然估计量 $\hat{p} = \frac{\bar{X}}{m}$.

$$5. \hat{\alpha} = -\frac{n}{\sum_{i=1}^n \ln x_i} - 1.$$

$$\hat{\alpha} = \frac{1}{1 - \bar{Z}}.$$

6. 置信区间

$$(\mu, \sigma^2) = \left(\bar{X} - z_{\alpha/2} \frac{\bar{\sigma}}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\bar{\sigma}}{\sqrt{n}} \right).$$

$$7. \left(14.98 - 1.96 \sqrt{\frac{0.25}{5}}, 14.98 + 1.96 \sqrt{\frac{0.05}{5}} \right) = (14.8, 15.2).$$

$$8. (432.3, 482.69).$$

$$9. (0.02, 0.10).$$

检测题(二)

1. 设总体 X 有概率密度

$$f(x; \theta) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 < x < \theta_2, \\ 0, & \text{其他.} \end{cases}$$

其中 $\theta_2 > \theta_1$ 为待估计参数, $x_1, x_2, \dots, x_n$ 为一组样本, 求 θ_1, θ_2 的矩估计量和极大似然估计.

2. 设总体 X 服从均匀分布, 其密度函数

$$f(x; \theta) = \begin{cases} 1, & x \in [\theta - 1/2, \theta + 1/2], \\ 0, & \text{其他.} \end{cases}$$

θ 为未知参数, $X_1, \dots, X_n$ 是一个样本, 试求 θ 的矩估计和极大似然估计量.

3. 设 (X_1, X_2) 是取自正态分布 $N(\mu, 1)$ 的一个容量为 2 的样本, 试证三个估计量都是 μ 的无偏估计量: $\hat{\mu}_1 = \frac{2}{3}X_1 + \frac{1}{3}X_2$, $\hat{\mu}_2 = \frac{1}{4}X_1 + \frac{3}{4}X_2$, $\hat{\mu} = \frac{1}{2}X_1 + \frac{1}{2}X_2$, 并指出其中哪一个方差最小.

4. 设 $(X_1, X_2, \dots, X_n)$ 是随机变量 X 的一个样本, 试证估计量 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $W = \sum_{i=1}^n a_i X_i$

($a_i \geq 0$ 为常数, $\sum_{i=1}^n a_i = 1$) 都是 EX 的无偏估计, 且 $\bar{X}$ 的方差不超过 W 的方差.

5. 假设总体 X 在区间 $[0, \theta]$ 上有均匀分布, 根据来自 X 的简单随机变量 $X_1, \dots, X_n$, 建立未知参数 θ 的 $1 - \alpha$ 置信区间.

6. 在一批货物的容量为 100 的样本中, 经检验发现有 16 只次品. 试求这批货物次品率的置信度为 0.95 的置信区间.

7. 对某一距离测量 5 次, 数据如下(单位: m):

2 781 2 836 2 807 2 763 2 858

认为测量值服从以距离真值为数学期望的正态分布,求距离真值的 95% 的置信区间.

8. 为了解灯泡使用时数的均值 μ 及标准差 σ ,测得 10 个灯泡,得 $\bar{x} = 1\,500$ h, $s = 20$ h. 如果已知灯泡的使用时数服从正态分布,求 μ 和 σ 的 95% 的置信区间.

检测题(二)参考答案

$$1. \begin{cases} \hat{\theta}_1 = \bar{X} - \sqrt{3}S, \\ \hat{\theta}_2 = \bar{X} + \sqrt{3}S \end{cases}$$

为 θ_1, θ_2 的矩估计量,其中 $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

2. $\theta = \bar{X}$ 为 θ 的矩估计.

因 $\left[X_n^* - \frac{1}{2}, X_1^* + \frac{1}{2} \right]$ 中任何值都是 θ 的极大似然估计,除非 $X_n^* - 1/2 = X_1^* + \frac{1}{2}$, 否则它是不唯一的.

3. $\hat{\mu}_3$ 的方差最小.

$$5. \left(\frac{X(n)}{\sqrt{\frac{n}{1-\alpha}}/2}, \frac{X(n)}{\sqrt{\frac{n}{\alpha}}/2} \right).$$

6. (0.101 0.244).

7. (2 760.78 2 857.22).

8. (1 485.69, 1 514.32).

(2)(13.8 36.5).

第7章 假设检验

7.1 假设检验

7.1.1 知识要点

1. 假设检验与参数检验

预先对总体分布函数的某些参数或分布函数的形式做某种假设,然后利用样本的有关信息,对所做假设的正确性进行推断,这类统计问题叫做假设检验,所做的假设叫做原假设(或统计假设),仅对总体分布函数中的参数提出假设进行检验的问题,叫做参数检验.

2. 统计思想

假设检验的统计思想是小概率原理,即概率很小的事件在一次试验中几乎是不可能发生的.先对总体的参数或分布函数的表达式做出某种假设,然后找出一个在假设成立条件下出现可能性甚小的小概率事件.如果试验或抽样的结果使该小概率事件发生了,这与小概率原理相违背,表明原来的假设有问题应予以否定,即拒绝接受这个假设.若该小概率事件在一次试验或抽样中并未发生,就没有理由否定原假设,这时称假设与试验结果是相容的.

3. 显著性水平与拒绝域

小概率原理关于小概率的值并没有统一规定,通常根据实际问题的要求,取 $\alpha = 0.1, 0.05, 0.01, 0.005, 0.001$ 等, α 称为检验中的显著性水平.

在假设检验过程中,使得小概率事件发生的统计量取值范围称为该假设检验的拒绝域,拒绝域的边界称为该检验的临界值.拒绝域的大小,依赖于显著性水平 α 的取值,一般说来,显著性水平越高,即 α 越小,拒绝域亦越小.

4. 假设检验的步骤

- ① 根据实际问题的要求,提出原假设 H_0 (也称零假设)及备择假设 H_1 ;
- ② 构造适当的统计量,并在原假设 H_0 成立的条件下确定该统计量的分布;
- ③ 给定显著性水平 α ;
- ④ 对给定的 α ,由 H_1 的内容确定拒绝域的形式,通常在水平 α 下,查相应统计量的分布表来决定临界值,由此确定检验的拒绝域;
- ⑤ 根据样本观测值算出统计量的观测值,若统计量的值落入拒绝域 W ,则拒绝原假设 H_0 ,否则接受 H_0 .

5. 两类错误

拒绝 H_0 要承担一定的风险,有可能将正确的假设误认为是错误的,在统计中称这种“以真为假”的错误为第Ⅰ类错误,即 H_0 实际上为真时,而拒绝 H_0 . 犯第Ⅰ类错误的概率显然是显著性水平 α ,即

$$P\{\text{拒绝 } H_0 \mid H_0 \text{ 为真}\} = \alpha.$$

由于显著性水平 α 一般取比较小的值,所以犯第Ⅰ类错误是小概率事件.

同样,假如 H_0 本来是假的,而接受 H_0 ,这就犯了“取伪”错误,称为第Ⅱ类错误. 犯第Ⅱ类错误的概率是

$$P\{\text{接受 } H_0 \mid H_0 \text{ 不真}\} = \beta.$$

7.1.2 典型题型解析

1. 假设检验统计量的辨识

假设检验是在对总体分布或参数做出假设的基础上,利用统计量来对假设做出判断,常见的是显著性检验. 因此选择何种统计量是假设检验的基础和前提. 总体、假设和检验目的决定了选取或构造何种统计量,这部分内容教材中一般都给出了明确的结果.

例 7.1.1 设总体 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \dots, X_n$ 是来自 X 的样本,记 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $Q^2 = \sum_{i=1}^n (X_i - \bar{X})^2$, 当 μ 和 σ^2 未知时,则

(1) 检验假设 $H_0: \mu = \mu_0$ 所使用的统计量是_____.

(2) 检验假设 $H_0: \sigma^2 = \sigma_0^2$ 使用的统计量是_____.

解 (1) 当 σ^2 未知时,检验假设 $H_0: \mu = \mu_0$, 应选服从 $n-1$ 个自由度的 t 分布的统计量

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

其中 $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ 为样本标准差,于是

$$t = \frac{\bar{X} - \mu_0}{\frac{Q}{\sqrt{n(n-1)}}}.$$

(2) 检验假设 $H_0: \sigma^2 = \sigma_0^2$, 应选统计量 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{Q^2}{\sigma_0^2}$.

2. 提出适当的假设与确定拒绝域形式

在进行假设检验时,首先要提出原假设 H_0 和备择假设 H_1 , 由于 H_0, H_1 的地位不是对等的, H_0 是受到保护的,因而选哪一个作为原假设需要小心. 一个原则是选择 H_0, H_1 使得两类错误中,后果严重的成为第Ⅰ类错误. 若两类错误中没有一类错误的后果严重需要避免时,常取 H_0 为维持现状,即取 H_0 为“无效益”、“无改进”、“无价值”等.

当写出 H_0, H_1 之后,拒绝域的形式是由备择假设所确定的. 拒绝域的形式直观上是可以

看得出来的. 例如采用 χ^2 检验法对正态总体检验假设 $H_0: \sigma^2 \geq \sigma_0^2, H_1: \sigma^2 < \sigma_0^2$, 这是左边检验, 当 H_1 为真时, $\frac{(n-1)S^2}{\sigma_0^2}$ 有偏小的倾向, 其拒绝域的形式为 $\frac{(n-1)S^2}{\sigma_0^2} \leq k$, 拒绝域为

$$\frac{(n-1)S^2}{\sigma_0^2} \leq \chi_{1-\alpha}^2(n-1).$$

例 7.1.2 设总体 $X \sim N(\mu, \sigma^2)$, 待检的原假设 $H_0: \sigma^2 = \sigma_0^2$, 对于给定的显著性水平 α , 如果拒绝域为 $(\chi_{\alpha}^2(n-1), +\infty)$, 则相应的备择假设 H_1 : _____, 若拒绝域为 $(0, \chi_{1-\frac{\alpha}{2}}^2(n-1)) \cup [\chi_{\frac{\alpha}{2}}^2(n-1), +\infty)$, 则相应的备择假设 H_1 : _____.

解 由题设, 原假设 $H_0: \sigma^2 = \sigma_0^2$. 选择 χ^2 分布的统计量 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$, 如果拒绝域为 $(\chi_{\alpha}^2(n-1), +\infty)$, 则属于单侧检验中的右边检验, 故对应的备择假设为 $H_1: \sigma^2 > \sigma_0^2$; 若拒绝域为 $(0, \chi_{1-\frac{\alpha}{2}}^2(n-1)) \cup [\chi_{\frac{\alpha}{2}}^2(n-1), +\infty)$, 属于双边检验, 备择假设 $H_1: \sigma^2 \neq \sigma_0^2$.

例 7.1.3 设总体 $X \sim N(\mu, \sigma^2)$, μ 和 σ^2 均未知, 统计假设取为 $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$, 若用 t 检验法进行假设检验, 则在显著水平 α 之下, 拒绝域是()

- A. $|t| < t_{\frac{\alpha}{2}}(n-1)$ B. $|t| \geq t_{\frac{\alpha}{2}}(n-1)$
C. $|t| \geq t_{\alpha}(n-1)$ D. $|t| < -t_{\alpha}(n-1)$

解 由题设可知, 此问题为双边检验问题, 依统计假设知, 应选 B.

3. 两类错误的辨识

若检验显著性水平为 α , 则“弃真”第 I 类错误的概率为 $P\{\text{拒绝 } H_0 | H_0 \text{ 为真}\} = \alpha$. 犯第 II 类错误即“取伪”的概率 $\beta = P\{\text{接受 } H_0 | H_0 \text{ 不真}\}$.

例 7.1.4 在假设检验中, 原假设 H_0 , 备择假设 H_1 , 则称()为犯第 II 类错误.

- A. H_0 为真, 接受 H_0 B. H_0 不真, 接受 H_0
C. H_0 为真, 拒绝 H_1 D. H_0 不真, 拒绝 H_0

解 按规定犯第 II 类错误, 就是犯“取伪”的错误, 即 $\{\text{接受 } H_0 | H_0 \text{ 不真}\}$, 故选 B.

例 7.1.5 某种产品以往的废品率为 5%, 采取某种技术革新措施后, 对产品的样本进行检验, 这种产品的废品率是否有所降低, 取显著性水平 $\alpha = 5\%$, 则此问题的零假设 H_0 : _____, 备择假设 H_1 : _____, 犯第 I 类错误的概率为 _____.

解 由题设可知, 零假设 $H_0: p = 5\%$, 备择假设 $H_1: p < 5\%$, 犯第 I 类错误的概率为 $\alpha = 5\%$.

同步练习 7.1

1. 关于零假设 H_0 的选取, 下列叙述错误的是().

- A. 尽量使后果严重的错误成为第 I 类错误
B. 可以根据检验结果随时改换 H_0 , 以达到希望得到的结论
C. 若拟从样本数据得到对某一结论强有力的支持, 则将此结论的对立面设为 H_0
D. 将不容易否定的诊断选作零假设

2. 关于检验水平 α 的设定, 下列叙述错误的是().
- A. α 的选取本质上是个实际问题, 而非数学问题
B. 在检验实施之前, α 应是事先给定的, 不可擅自改动
C. α 即为检验结果犯第 I 类错误的最大概率
D. 为了得到所希望的结论, 可随时对 α 的值进行修正
3. 关于对检验结果的处理, 下列叙述不正确的是().
- A. 若样本落入拒绝域, 由于 α 值很小, 则否定零假设是很有把握的
B. 若样本未落入拒绝域, 则可以放心地接受 H_0
C. 若样本未落入拒绝域, 同时 β 的值很小时, 可以以较大的把握接受 H_0
D. 可以用检验的 P 值来辅助判断
4. 下列关于“拒绝域”的评述中, 不正确的是().
- A. 拒绝域是样本空间(即全体样本点的集合)的子集
B. 拒绝域的结构形式是先定的, 与具体抽样结果无关
C. 拒绝域往往是通过某检验统计量诱导出来的
D. 拒绝域中涉及的临界值要通过抽样来确定
5. 进行假设检验时, 若增大样本容量, 其他条件不变, 则犯两类错误的概率().
- A. 都增大
B. 都减小
C. 都不变
D. 一个增大, 一个减小
6. 在假设检验中, 一般情况下().
- A. 只犯第 I 类错误
B. 只犯第 II 类错误
C. 两类错误都可能发生
D. 不会犯错误

同步练习 7.1 参考答案

1. B 2. D 3. B 4. D 5. B 6. C

7.2 单个正态总体的假设检验

7.2.1 知识要点

1. 单个正态总体的假设检验的类型

正态总体 $N(\mu, \sigma^2)$ 的假设检验问题, 主要有以下几种:

- ① 已知方差 σ^2 , 关于 μ 的检验;
- ② 未知方差 σ^2 , 关于 μ 的检验;
- ③ 已知均值 μ , 关于 σ 的检验;
- ④ 未知均值 μ , 关于 σ 的检验.

2. 单个正态总体均值、方差的假设检验

单个正态总体均值、方差检验法(显著性水平为 α)

检验参数	条件	假设		检验统计量	H_0 为真时分布	拒绝域
		H_0	H_1			
μ	σ^2 已知 $\sigma = \sigma_0$	$\mu = \mu_0$	$\mu \neq \mu_0$	$U = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}}$ 其中 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$N(0, 1)$	$ u \geq z_{\frac{\alpha}{2}}$
		$\mu \leq \mu_0$	$\mu > \mu_0$			$u \geq z_\alpha$
		$\mu \geq \mu_0$	$\mu < \mu_0$			$u \leq -z_\alpha$
	σ^2 未知	$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$ 其中 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$	$t(n-1)$	$ t \geq t_{\frac{\alpha}{2}}(n-1)$
		$\mu \leq \mu_0$	$\mu > \mu_0$			$t \geq t_\alpha(n-1)$
		$\mu \geq \mu_0$	$\mu < \mu_0$			$t \leq -t_\alpha(n-1)$
σ^2	μ 未知	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2(n-1)$	$\chi^2 \geq \chi_{\alpha/2}^2(n-1)$ 或 $\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)$
		$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$			$\chi^2 \geq \chi_\alpha^2(n-1)$
		$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$			$\chi^2 \leq \chi_{1-\alpha}^2(n-1)$
	μ 已知 $\mu = \mu_0$	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu_0)^2$	$\chi^2(n)$	$\chi^2 \geq \chi_{\frac{\alpha}{2}}^2(n)$ 或 $\chi^2 \leq \chi_{1-\alpha/2}^2(n)$
		$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$			$\chi^2 \geq \chi_\alpha^2(n)$
		$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$			$\chi^2 \leq \chi_{1-\alpha}^2(n)$

7.2.2 典型题型解析

1. 已知方差 σ^2 均值 μ 的检验

设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, 检验统计量 $U = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} \sim N(0, 1)$.

若 $H_0: \mu = \mu_0$, 则拒绝域为 $|u| \geq z_{\frac{\alpha}{2}}$; 若 $H_0: \mu \leq \mu_0$, 则拒绝域 $u \geq z_\alpha$; 若 $H_0: \mu \geq \mu_0$, 则拒绝域 $u \leq -z_\alpha$.

例 7.2.1 某切割机在正常工作时, 切割每段金属棒的平均长度为 10.5 cm. 设切割机切割每段金属棒的长度服从正态分布, 且根据长期经验知其方差 $\sigma_0^2 = 0.15^2 \text{ cm}^2$. 某日为了检验切割机工作是否正常, 随机地抽取 15 段进行测量, 其结果如下(单位: cm):

10.4, 10.6, 10.1, 10.4, 10.5, 10.3, 10.3, 10.2, 10.9, 10.6, 10.8, 10.5, 10.7, 10.2, 10.7

试问该机工作是否正常($\alpha = 0.05$)?

解 $H_0: \mu = \mu_0 = 10.5$, $H_1: \mu \neq 10.5$.

利用统计量 $U = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}}$ 检验, 在 H_0 为真时, $U \sim N(0, 1)$. 因 $\alpha = 0.05$, 由正态 $N(0, 1)$ 表查得分位点 $z_{0.025} = 1.96$.

由所给数据算得

$$|u| = \left| \frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n} \right| = \left| \frac{1.48 - 10.5}{0.15} \sqrt{15} \right| = 0.52.$$

因 $|u| = 0.52 < 1.96$, 故不能拒绝零假设 H_0 , 因而认为该切割机工作正常.

例 7.2.2 要求一种元件使用寿命不得低于 1 000 h, 今从一批这种元件中随机抽取 25 个, 测得其寿命的平均值为 950 h. 已知该种元件寿命服从标准差为 $\sigma = 100$ h 的正态分布. 试在显著性水平 $\alpha = 0.05$ 下确定这批元件是否合格?

解 据题意, 检验的假设为

$$H_0: \mu \geq 1\,000, H_1: \mu < 1\,000.$$

检验统计量 $U = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}}$ 在 H_0 为真时服从 $N(0, 1)$. $\alpha = 0.05$, 查 $N(0, 1)$ 表得 $z_\alpha = z_{0.05} = 1.645$.

$$u = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} = \frac{950 - 1\,000}{100 / \sqrt{25}} = -2.5.$$

本检验为单边检验, 拒绝域为 $u \leq -z_\alpha$.

$$\text{由 } u = -2.5 < -1.645 = -z_\alpha$$

所以拒绝 H_0 , 即认为这批元件不合格.

例 7.2.3 某厂对废水进行处理, 要求某种有毒物质的浓度小于 19 (kg/L). 抽样检查得到 10 个数据, 其样本均值 $\bar{x} = 17.1$ (kg/L). 设有毒物质的含量服从正态分布, 且已知方差 $\sigma^2 = 8.5$ (kg²/L²). 问在显著水平 $\alpha = 0.05$ 下, 处理后的废水是否合格?

解 由题意, 检验的假设为

$$H_0: \mu = 19, H_1: \mu < 19.$$

由 σ^2 已知, 故检验统计量 $U = \frac{\bar{X} - \mu}{\sigma_0 / \sqrt{n}} \sim N(0, 1)$.

$\alpha = 0.05$, 查表 $z_\alpha = z_{0.05} = 1.645$, 拒绝域 $u \leq -z_\alpha$.

$$u = \frac{17.1 - 19}{\sqrt{8.5} / \sqrt{10}} = -2.06 < -1.64,$$

所以拒绝 H_0 , 即认为 $\mu < 19$, 处理后的水合格.

注 对于一个实际问题, 如何选择原假设 H_0 和备择假设 H_1 , 要根据问题的具体目的和要求而定. 例如, 我们要检验一些产品是否合格, 且希望在产品合格的情况下判断为不合格的概率得到控制, 就应取“产品合格”为 H_0 , “产品不合格”为 H_1 . 若问题是要决定新提方法(新材料、新工艺、新配方等)是否比原方法好, 往往将“原方法”取为 H_0 , 而将“新方法”取为 H_1 .

例 7.2.4 设 $X_1, X_2, \dots, X_{16}$ 为来自总体 $N(\mu, 4)$ 的样本, μ 未知, $\bar{X}$ 为样本均值, 如果对检验问题 $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$, 取拒绝域 $W = \{|\bar{X} - \mu_0| \geq k\}$. 试决定 k , 检验的显著水平为 0.05.

解 由于 $X \sim N(\mu, 4)$, 在 H_0 为真时 $\bar{X} \sim N\left(\mu_0, \frac{4}{16}\right)$.

$$P(|\bar{X} - \mu_0| \geq k) = P\left(\left|\frac{\bar{X} - \mu_0}{\sqrt{\frac{4}{16}}}\right| \geq k \sqrt{\frac{16}{4}}\right) = 0.05,$$

因为 $\frac{\bar{X} - \mu_0}{\sqrt{\frac{4}{16}}} \sim N(0, 1)$, 由正态表查得 $z_{\frac{\alpha}{2}} = 1.96$, 即

$$k\sqrt{\frac{16}{4}} = 1.96, \text{ 所以 } k = \frac{1.96}{2} = 0.98.$$

例 7.2.5 设总体 $X \sim N(\mu, \sigma^2)$, 在 $\alpha = 0.05$ 的水平上检验

$$H_0: \mu = 0, H_1: \mu \neq 0.$$

如果所选取的拒绝域 $W = \{|\bar{X}| \geq 1.96\}$, 问样本容量 n 应取多大.

解 由题意 $P\{|\bar{X}| \geq 1.96\} = 0.05$,

$$P\{|\bar{X}| < 1.96\} = P\left\{\left|\frac{\bar{X}}{S}\sqrt{n}\right| < 0.392\sqrt{n}\right\} = 0.95.$$

由于 $\frac{\bar{X}}{S}\sqrt{n} \sim N(0, 1)$, 因此, 有

$$0.392\sqrt{n} = 1.96, \text{ 故 } n = 25.$$

2. σ^2 未知, 关于 μ 的检验

设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, σ^2 未知时, 检验统计量 $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$.

若 $H_0: \mu = \mu_0$, 则拒绝域 $|t| \geq t_{\frac{\alpha}{2}}(n-1)$; 若 $H_0: \mu \leq \mu_0$, 则 $t \geq t_{\alpha}(n-1)$; 若 $H_0: \mu \geq \mu_0$, 则 $t \leq -t_{\alpha}(n-1)$. 解题步骤为 ①提出假设; ②选取统计量, 并指出分布; ③求出拒绝域; ④计算统计量, 做出判断.

例 7.2.6 某批矿砂的 5 个样品中的镍含量经测定为(%):

$$3.25, 3.27, 3.24, 3.26, 3.24.$$

设测定值总体服从正态分布. 问在 $\alpha = 0.01$ 下能否接受假设: 这批矿砂的镍含量的均值为 3.25.

解 要在水平 $\alpha = 0.01$ 下检验假设:

$$H_0: \mu = \mu_0 = 3.25, H_1: \mu \neq \mu_0.$$

由于总体 $N(\mu, \sigma^2)$ 的方差 σ^2 未知, 故可用 t 检验法. 先算得

$$\bar{x} = 3.252, s = 0.0130.$$

由样本值算得统计量 T 的值为

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.002}{0.013} \times \sqrt{5} = 0.344.$$

对给定 $\alpha = 0.01$, 查 $t(n-1)$ 表得 $t_{\frac{\alpha}{2}}(4) = t_{0.005}(4) = 4.6041$, 由 $|t| = 0.344 < 4.6041$, 故接受 H_0 , 即认为这批矿砂的镍含量为 3.25.

例 7.2.7 用某种钢生产的钢筋的强度 X 服从正态分布, 且 $E(X) = 50.00(\text{kg/mm}^2)$. 今改变炼钢的配方, 利用新法炼了 9 炉钢, 从这 9 炉钢生产的钢筋中每炉抽一根测得其强度分别为:

$$56.01, 52.45, 51.53, 48.52, 49.04, 53.38, 54.02, 52.13, 52.15.$$

试问用新法炼钢生产的钢筋其强度的均值是否有明显提高($\alpha=0.05$)?

解 根据问题的特点,假设为

$$H_0: \mu = 50.00, H_1: \mu > 50.00.$$

由于总体的方差未知,所以选用统计量 T 在 H_0 为真下,

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1),$$

对显著性水平 $\alpha=0.05$ 查表得 $t_{0.05}(8)=1.86$ 因而拒绝域为 $W=[1.86, +\infty)$.

由样本观察值算得 $\bar{x}=52.14, s=2.24$,

$$t = \frac{52.14 - 50.00}{2.24/\sqrt{9}} = 2.87 > t_{0.05}(8) = 1.86.$$

所以拒绝 H_0 ,即认为强度有明显提高.

例 7.2.8 按规定,每 100 g 的罐头,番茄汁中 V_c (维生素 C)的含量不少于 21 mg.现从某厂生产的一批罐头中抽取 17 个,得 V_c 的含量(单位:mg)为 16, 22, 21, 20, 23, 21, 19, 15, 13, 23, 17, 20, 29, 18, 22, 16, 25. 已知 V_c 的含量服从正态分布,试以 0.025 的检验水平检验该批罐头的 V_c 的含量是否合格.

解 由题设待检假设为

$$H_0: \mu \geq 21, H_1: \mu < 21.$$

σ^2 未知,检验统计量 $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$,对 $\alpha=0.025$ 查 t 分布表

$$t_{\alpha}(16) = t_{0.025}(16) = 2.12.$$

由样本观察值算得 $\bar{x} = \frac{1}{17} \times 340 = 20, S = 3.87$,

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{20 - 21}{3.87/\sqrt{17}} = -1.065.$$

由 $t = -1.065 > -2.12 = -t_{0.025}(16)$,

故接受 H_0 ,即可认为该批罐头的 V_c 含量是合格的.

3. 均值 μ 未知,方差 σ^2 的检验

总体为 $N(\mu, \sigma^2)$, μ, σ^2 未知, $X_1, X_2, \dots, X_n$ 为样本,则统计量 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$. $H_0: \sigma^2 = \sigma_0^2$, 则拒绝域为 $\chi^2 \geq \chi_{\alpha/2}^2(n-1)$ 或 $\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)$; $H_0: \sigma^2 \leq \sigma_0^2$, 则拒绝域为 $\chi^2 \geq \chi_{\alpha}^2(n-1)$; $H_0: \sigma^2 \geq \sigma_0^2$, 则拒绝域为 $\chi^2 \leq \chi_{1-\alpha}^2(n-1)$.

例 7.2.9 一细纱车间纺出某种细纱支数标准差为 1.2,某日从纺出的一批纱中,随机抽 15 缕进行支数测量,测得子样标准差 $s=2.1$,问纱的均匀度有无显著变化($\alpha=0.05$),假定总体分布是正态分布.

解 该日纺出纱的支数构成一个正态总体,按题意要检验假设为

$$H_0: \sigma^2 = 1.2^2, H_1: \sigma^2 \neq 1.2^2.$$

由于 μ 未知, 统计量 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$.

由 $\alpha = 0.05$, 自由度 $n-1 = 14$, 查 χ^2 分布表得 $\chi_{\frac{\alpha}{2}}^2(n-1) = 26.119$, $\chi_{1-\frac{\alpha}{2}}^2(n-1) = 5.629$.

由样本观察值算得 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = 14 \times \frac{2.1^2}{1.2^2} = 42.875$.

易见 $\chi^2 = 42.875 > 26.119 = \chi_{\frac{\alpha}{2}}^2(n-1)$, 所以拒绝假设 H_0 , 即这天细纱均匀度有显著变化.

例 7.2.10 某种导线, 要求其电阻的标准差不得超过 0.005Ω . 今在生产的一批导线中取样品 9 根, 测得 $s = 0.007 \Omega$. 问在显著性水平 $\alpha = 0.05$ 条件下, 能认为这批导线的方差显著偏大吗?

解 $H_0: \sigma^2 = (0.005)^2$, $H_1: \sigma^2 > (0.005)^2$.

由于 μ 未知, 取统计量 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$.

由 $\alpha = 0.05$, 查表 $\chi_{\alpha}^2(8) = \chi_{0.05}^2(8) = 15.5$, 拒绝域为 $\chi^2 \geq \chi_{0.05}^2(8)$. 由样本观察值计算得

$$\chi^2 = \frac{8 \times 0.007^2}{0.005^2} = 15.68 > 15.5.$$

故拒绝 H_0 , 认为这批导线的方差显著地偏大.

例 7.2.11 已知某厂生产的滚珠直径服从正态分布, 按照规定平均直径为 5 mm , 标准差不超过 0.5 mm , 现从这批滚珠中随机抽取 9 个, 测得其平均直径为 4.78 mm , 标准差为 0.6 mm , 问这批滚珠的质量是否合格 ($\alpha = 0.05$).

解 滚珠质量合格有两个标准, 一是均值为 5 mm , 二是方差不超过 $0.5^2 (\text{mm})^2$, 因此这是一个正态总体下两个假设检验问题.

(1) $H_0: \mu = 5$, $H_1: \mu \neq 5$.

σ^2 未知, 检验统计量 $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$.

由 $\alpha = 0.05$, 查表 $t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(8) = 2.31$, 拒绝域为 $|t| \geq 2.31$.

$$t = \frac{4.78 - 5}{0.6/\sqrt{9}} = -1.1,$$

$|t| = 1.1 < 2.31$, 因此不能拒绝 H_0 , 即可以认为滚珠的平均直径符合规定.

(2) $H'_0: \sigma^2 \leq 0.5^2$, $H'_1: \sigma^2 > 0.5^2$.

μ 未知, 在 H'_0 为真下, $\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,

$\alpha = 0.05$, 查表得 $\chi_{0.05}^2(8) = 15.51$, 拒绝域为 $\chi^2 \geq 15.51$, 由观察值算得

$$\chi^2 = \frac{8 \times 0.6^2}{0.5^2} = 11.52 < 15.51.$$

所以不能拒绝 H'_0 , 即认为标准差符合规定.

综合对滚珠的平均直径和方差的检验结果, 可以认为这批滚珠合格.

4. 均值 μ 已知, 方差 σ^2 的检验

此时检验统计量 $\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu_0)^2 \sim \chi^2(n)$, 拒绝域与 μ 未知时的类似.

例 7.2.12 某厂管理部门要求一种产品的重量方差不得超过 30, 现随机抽取 8 件产品检测, 算得 $\sum_{i=1}^8 x_i = 21$, $\sum_{i=1}^8 x_i^2 = 513$, 由经验知产品的重量服从正态分布, 其中 $\mu_0 = 1$, 问产品是否符合标准 ($\alpha = 0.05$).

解 由题意分析, 待检假设为

$$H_0: \sigma^2 \leq 30, H_1: \sigma^2 > 30.$$

由于 $\mu = \mu_0 = 1$ 已知, 所以用统计量

$$\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu_0)^2 \sim \chi^2(8) (\sigma^2 \leq 30 \text{ 为真情况下}).$$

$\alpha = 0.05$, $\chi_{\alpha}^2 = \chi_{0.05}^2(8) = 15.507$, 拒绝域为 $\chi^2 \geq \chi_{\alpha}^2(8) = 15.507$,

而 $\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu_0)^2 = \frac{1}{\sigma_0^2} (\sum_{i=1}^n X_i^2 - 2\mu_0 \sum_{i=1}^n X_i + n\mu_0^2)$

$$= \frac{1}{30} (513 - 2 \times 21 \times 8) = 16 > 15.507.$$

故拒绝 H_0 , 即认为产品重量的方差超过 30, 不合标准.

同步练习 7.2

1. 设 $X_1, X_2, \dots, X_n$ 为取自总体 $N(\mu, \sigma^2)$ 的样本, 参数 μ, σ^2 均未知, 且 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$,

$Q^2 = \sum_{i=1}^n (X_i - \bar{X})^2$, 则假设 $H_0: \mu = 0$ 的检验使用的统计量为_____.

2. 设样本 $X_1, X_2, \dots, X_n$ 抽自总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2 均未知. 要对 μ 进行假设检验 $H_0: \mu = \mu_0$ (μ_0 已知), $H_1: \mu \neq \mu_0$, 则要用的检验统计量为_____, 给定显著水平 α , 则检验的拒绝域为_____, 若对 μ 进行假设检验 $H_0: \mu = \mu_0, H_1: \mu < \mu_0$, 则检验的拒绝域为_____.

3. 设样本 $X_1, X_2, \dots, X_n$ 来自总体 $X \sim N(\mu, \sigma^2)$, μ 已知, 要对 σ^2 做假设检验, 统计假设为 $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$, 则检验统计量为_____, 给定显著水平 α , 则检验的拒绝域为_____. 若统计假设为 $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 > \sigma_0^2$, 则检验的拒绝域为_____.

4. 长期统计资料表明, 某市轻工产品月产值的百分比 X 服从正态分布, 方差 $\sigma^2 = 1.21$. 现任意抽查 10 个月, 得轻工产品产值占总产值的百分比为 31.31%, 30.10%, 32.16%, 32.56%, 29.66%, 31.64%, 31.64%, 30.00%, 31.87%, 31.03%. 问在置信度为 5% 时, 可否认为过去该市轻工产品月产值占该市工业产品总月产值百分比的平均数为 32.50%.

5. 某装置的平均工作温度据制造厂家称不高于 190°C . 从一个由 16 台装置构成的随机样本测得工作温度的平均值和标准差分别为 195°C 和 8°C . 根据这些数据能否说明平均工作温度比制造厂所说的要高? 设 $\alpha = 0.05$, 并假定工作温度近似服从正态分布.

6. 正常人的脉搏平均为 72 次/分, 医生测得 10 例慢性四乙基铅中毒患者的脉搏(次/分): 54, 67, 68, 78, 70, 66, 67, 70, 65, 69. 已知脉搏服从正态分布, 问在显著性水平 $\alpha = 0.05$ 条件下, 四乙基铅中毒者和正常人的脉搏有无显著性差异?

7. 由某个正态总体中抽出一个容量为 21 的简单随机样本, 得修正样本方差为 10, 能否根据此结果得出总体方差小于 15 的结论 ($\alpha = 0.05$)?

8. 用包装机包装洗衣粉, 在正常的情况下, 每袋标准重量为 1 000 g, 标准差 σ 不能超过 15 g. 假设洗衣粉袋重服从正态分布, 某天检验包装机工作情况, 从已装好的袋中随机抽取 10 袋, 测得其重(单位: g)为 1 020, 1 030, 968, 994, 1 014, 998, 976, 982, 950, 1 048, 问按标准差来衡量, 这天机器工作是否正常 ($\alpha = 0.05$)?

同步练习 7.2 参考答案

$$1. \frac{\bar{X} \sqrt{n(n-1)}}{Q}.$$

$$2. t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad (-\infty, -t_{\frac{\alpha}{2}}(n-1)) \cup [t_{\frac{\alpha}{2}}(n-1), +\infty) \quad (-\infty, -t_{\alpha}(n-1)]$$

$$3. \chi^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} \quad (0, \chi_{1-\alpha/2}^2(n)) \cup [\chi_{\frac{\alpha}{2}}^2(n), +\infty) \quad [\chi_{\alpha}^2(n), +\infty).$$

$$4. u = -3.7427, \text{拒绝域为 } |u| \geq 1.96, \text{所以拒绝 } H_0.$$

5. $H_0: \mu \leq 190$, $H_1: \mu > 190$, 查表 $t_{0.05}(15) = 1.75$, 由 $t = 2.5 > 1.75$, 拒绝 H_0 , 即平均工作温度比制造厂所说的要高.

6. $H_0: \mu = \mu_0 = 72$, $H_1: \mu \neq 72$, 查表 $t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(9) = 2.262$, $t = 2.45 > 2.262$, 拒绝 H_0 , 即认为四乙基铅中毒患者和正常人的脉搏有显著差异.

7. $H_0: \sigma^2 \geq 15$, $H_1: \sigma^2 < 15$. 查表 $\chi_{1-\alpha}^2(n-1) = \chi_{0.95}^2(20) = 10.851$, $\chi^2 = 13.333 > 10.851$, 所以不能否定 H_0 , 即不能得出总体方差小于 15 的结论.

8. $H_0: \sigma^2 \leq 15^2$, $H_1: \sigma^2 > 15^2$, $\chi_{0.05}^2(9) = 16.919$ (查表), $\chi^2 = 36.554 > 16.919$. 故拒绝 H_0 , 即认为这天包装机工作不正常.

7.3 两个正态总体的假设检验

7.3.1 知识要点

1. 两个正态总体假设检验的常见问题

设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 且来自 X 的样本 $X_1, X_2, \dots, X_{n_1}$ 和来自 Y 的样本 $Y_1, Y_2, \dots, Y_{n_2}$ 相互独立, 常见的问题有

(1) 已知 σ_1^2, σ_2^2 , 检验假设 μ_1, μ_2 关系;

(2) 未知 σ_1^2, σ_2^2 , 但已知 $\sigma_1^2 = \sigma_2^2$, 检验假设 μ_1, μ_2 关系;

(3) 未知 μ_1, μ_2 , 检验假设 σ_1^2, σ_2^2 关系.

2. 两个正态总体假设检验方法

检验参数	条件	假设		检验统计量	H_0 为真时分布	拒绝域
		原假设 H_0	备择假设 H_1			
μ_1, μ_2	已知 σ_1^2, σ_2^2	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$N(0, 1)$	$ u \geq z_{\frac{\alpha}{2}}$
		$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$			$u \geq z_{\alpha}$
		$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$			$u \leq -z_{\alpha}$
	σ_1^2, σ_2^2 未知, 但知 $\sigma_1^2 = \sigma_2^2$	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$t = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	$t(n_1 + n_2 - 2)$	$ t \geq t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)$
		$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$			$t \geq t_{\alpha}(n_1 + n_2 - 2)$
		$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$			$t \leq -t_{\alpha}(n_1 + n_2 - 2)$
σ_1^2, σ_2^2	μ_1, μ_2 未知	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$	$F(n_1 - 1, n_2 - 1)$	$F \geq F_{\frac{\alpha}{2}} \text{ 或 } F \leq F_{1 - \frac{\alpha}{2}}$
		$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$			$F \geq F_{\alpha}$
		$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$			$F \leq F_{1 - \alpha}$

3. 成对数据的比较

设有两个总体 X 和 Y , X 和 Y 不独立, $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ 为来自 (X, Y) 的样本, 记

$$d_i = x_i - y_i (i = 1, 2, \dots, n),$$
$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i, S^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2,$$

检验统计量 $T = \frac{\bar{d}}{S/\sqrt{n}} \sim t(n-1).$

成对数据 (X, Y) 均值 μ_1, μ_2 的检验归纳为下表：

检验目的	条件	假设		检验统计量	H_0 为真时分布	拒绝域
		原假设 H_0	备择假设 H_1			
比较成对数据的差异性	成对数据已知	$\mu_1 - \mu_2 = 0$ 或 $\mu_d = 0$	$\mu_1 - \mu_2 \neq 0$ 或 $\mu_d \neq 0$	$t = \frac{\bar{d}}{S/\sqrt{n}}$	$t(n-1)$	$ t \geq t_{\frac{\alpha}{2}}(n-1)$
		$\mu_1 - \mu_2 \leq 0$ 或 $\mu_d \leq 0$	$\mu_1 - \mu_2 > 0$ 或 $\mu_d > 0$			$t \geq t_{\alpha}(n-1)$
		$\mu_1 - \mu_2 \geq 0$ 或 $\mu_d \geq 0$	$\mu_1 - \mu_2 < 0$ 或 $\mu_d < 0$			$t \leq -t_{\alpha}(n-1)$

7.3.2 典型题型解析

1. 两正态总体的 u 检验

两正态总体 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2), X, Y$ 相互独立, σ_1^2, σ_2^2 已知, 检验 μ_1, μ_2 的

关系需要用 u 检验统计量,称为 u 检验. 解题步骤为:①提出假设;②计算统计量 $u =$

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; \text{③求临界值(可查标准正态分布表)}; \text{④做出检验结论.}$$

例 7.3.1 对市场上出售的两种同类食品,要检测它们的某种营养元素含量是否相同.从产品中各抽取质量相同的 5 例进行化验,测得其营养元素含量(单位:mg)为:

A 21 24 26 27 24

B 23 27 31 26 28

据经验知,这种营养元素含量服从正态分布,且 A, B 的方差分别为 5 和 8. 问 A, B 两种食品的营养元素含量有否显著差异($\alpha = 0.05$).

解 设 A, B 两种食品的营养元素含量分别为 X, Y , 其均值为 μ_1, μ_2 , 由题设条件得 $X \sim N(\mu_1, 5), Y \sim N(\mu_2, 8)$, 显然 X, Y 相互独立, 要检验 $\mu_1 = \mu_2$, 用两总体 u 检验法.

$$H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2.$$

$$\text{计算 } u \text{ 值: } \bar{x} = \frac{1}{5}(21 + 24 + 26 + 27 + 24) = 24.4,$$

$$\bar{y} = \frac{1}{5}(23 + 27 + 31 + 26 + 28) = 27,$$

$$n_1 = n_2 = 5, \sigma_1^2 = 5, \sigma_2^2 = 8, \text{ 这样}$$

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{24.4 - 27}{\sqrt{\frac{5}{5} + \frac{8}{5}}} = -1.612.$$

对给定的检验水平 $\alpha = 0.05$, 查标准正态分布表得临界值 $z_{\frac{\alpha}{2}} = 1.96$.

由于 $|u| = 1.612 < 1.96 = z_{\frac{\alpha}{2}}$, 所以接受 H_0 , 即认为 A, B 两种食品的某种营养元素含量并无显著差异.

例 7.3.2 一药厂生产一种新的止痛片, 厂方希望验证服用新药片后至开始起作用的时间间隔较原有止痛片至少缩短一半, 因此厂方提出需检验假设

$$H_0: \mu_1 \leq 2\mu_2, H_1: \mu_1 > 2\mu_2.$$

此处 μ_1, μ_2 分别是服用原有止痛片和服用新止痛片后至起作用的时间间隔总体的均值. 设两总体均为正态, 且方差分别为已知值 σ_1^2, σ_2^2 . 现分别在两总体中取一样本 $X_1, X_2, \dots, X_{n_1}$ 和 $Y_1, Y_2, \dots, Y_{n_2}$, 设两个样本独立. 试给出上述假设 H_0 的拒绝域. 取显著性水平为 α .

解 已知 $X_i \sim N(\mu_1, \sigma_1^2), i = 1, 2, \dots, n_1,$

$$Y_i \sim N(\mu_2, \sigma_2^2), i = 1, 2, \dots, n_2,$$

且两样本 $X_1, X_2, \dots, X_{n_1}$ 和 $Y_1, Y_2, \dots, Y_{n_2}$ 独立. 记

$$\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i, \bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i.$$

$$\text{检验统计量 } U = \frac{\bar{X} - 2\bar{Y}}{\sqrt{\sigma_1^2/n_1 + 4\sigma_2^2/n_2}}, H_0 \text{ 为真时}$$

$$U \sim N(0, 1).$$

对于显著性水平 α , 记临界值为 2α , 检验的拒绝域为

$$\frac{\bar{x} - 2\bar{y}}{\sqrt{\sigma_1^2/n_1 + 4\sigma_2^2/n_2}} \geq z_\alpha.$$

2. 两正态总体的 t 检验

两正态总体 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, X, Y 相互独立, σ_1^2, σ_2^2 未知, 但知 $\sigma_1^2 = \sigma_2^2$. 检验 μ_1, μ_2 的关系需要用 t 为检验统计量, 称为 t 检验. 解题步骤为: ①提出假设; ②计算统计量 $t = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, 其中 $S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$; ③确定临界值; ④做出检验结论.

例 7.3.3 根据以往的经验, 元件的电阻服从正态分布. 现在对 A, B 两批同类无线电元件的电阻进行测试, 测得如下结果(单位 Ω):

A 0.140 0.138 0.143 0.141 0.144 0.137

B 0.135 0.140 0.142 0.136 0.138 0.140

已知 $\sigma_1^2 = \sigma_2^2$, 能否认为两批元件的电阻无显著差异($\alpha = 0.05$).

解 $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$.

$$\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i = \frac{1}{6} (0.140 + 0.138 + \dots + 0.137) = 0.1405,$$

$$\bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i = \frac{1}{6} (0.135 + 0.140 + \dots + 0.140) = 0.1385,$$

$$(n_1 - 1)s_1^2 = \sum_{i=1}^{n_1} (x_i - \bar{x})^2 = 3.75 \times 10^{-5},$$

$$(n_2 - 1)s_2^2 = \sum_{i=1}^{n_2} (y_i - \bar{y})^2 = 3.55 \times 10^{-5},$$

$$t = \frac{\bar{x} - \bar{y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.1405 - 0.1385}{\sqrt{3.75 \times 10^{-5} + 3.55 \times 10^{-5}}} \sqrt{30} = 1.28.$$

对 $\alpha = 0.05$, 自由度 $n_1 + n_2 - 2 = 6 + 6 - 2 = 10$, 查表得

$$t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) = t_{0.025}(10) = 2.228.$$

由于 $|t| = 1.28 < 2.228$, 故接受 H_0 , 即认为两批元件的电阻无显著差异.

例 7.3.4 某灯泡厂在采用一项新工艺的前后, 分别抽取 10 个灯泡进行寿命试验. 计算得到: 采用新工艺前灯泡寿命的样本均值为 2460(h), 样本标准差为 56(h); 采用新工艺后灯泡寿命的样本均值为 2550(h), 样本标准差为 48(h). 设灯泡的寿命服从正态分布, 是否可以认为采用新工艺后灯泡的平均寿命有显著提高($\alpha = 0.01$)?

解 设采用新工艺前灯泡的寿命 $X \sim N(\mu_1, \sigma_1^2)$, 采用新工艺后灯泡的寿命 $Y \sim N(\mu_2, \sigma_2^2)$, σ_1^2, σ_2^2 未知, 假定 $\sigma_1^2 = \sigma_2^2$. 检验下面假设

$$H_0: \mu_1 = \mu_2, H_1: \mu_1 < \mu_2.$$

已知 $n_1 = n_2 = 10$, $\bar{x} = 2460$, $\bar{y} = 2550$, $s_1 = 56$, $s_2 = 48$.

$$S_w = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{9 \times 56^2 + 9 \times 48^2}{18}} \doteq 52.15,$$

$$t = \frac{\bar{x} - \bar{y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \doteq \frac{2\,460 - 2\,550}{52.15 \sqrt{\frac{1}{10} + \frac{1}{10}}} \doteq -3.86.$$

查表得临界值 $t_{\alpha}(n_1 + n_2 - 2) = t_{0.01}(18) = 2.55$.

由 $t < -t_{0.01}(18)$ 所以拒绝原假设 H_0 而接受备择假设 H_1 , 即认为采用新工艺后灯泡的平均寿命显著提高了.

3. 两正态总体的 F 检验

设两正态总体 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, X, Y 相互独立, μ_1, μ_2 未知, 检验 σ_1^2, σ_2^2 的关系需要用 F 统计量, 称为 F 检验. 解题步骤为: ①提出假设; ②计算统计量 $F = \frac{S_1^2}{S_2^2}$ (或 $F = \frac{\max\{S_1^2, S_2^2\}}{\min\{S_1^2, S_2^2\}}$); ③求临界值; ④做出检验结论.

例 7.3.5 甲、乙两台机床加工同一种轴. 从这两台机床加工的轴中分别随机地抽取若干根, 测得直径(单位 mm)为:

机床甲 20.5 19.8 19.7 20.4 20.1 20.0 19.0 19.9

机床乙 19.7 20.8 20.5 19.8 19.4 20.6 19.2

假设各台机床加工轴的直径分别构成正态总体. 试比较甲、乙二台机床加工的精度有无显著差异($\alpha = 0.05$).

解 按题意 本题是检验两正态总体 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ 方差是否相等的问题.

$$H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2.$$

$$n_1 = 8, \bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i = \frac{1}{8} (20.5 + 19.8 + \dots + 19.9) = 19.93,$$

$$\begin{aligned} s_1^2 &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 \\ &= \frac{1}{7} [(20.5 - 19.93)^2 + (19.8 - 19.93)^2 + \dots + (19.9 - 19.93)^2] = 0.216, \end{aligned}$$

$$n_2 = 7, \bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i = \frac{1}{7} (19.7 + 20.8 + \dots + 19.2) = 20.00,$$

$$\begin{aligned} s_2^2 &= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2 \\ &= \frac{1}{6} [(19.7 - 20.00)^2 + (20.8 - 20.00)^2 + \dots + (19.2 - 20.00)^2] = 0.397. \end{aligned}$$

$$F = \frac{S_1^2}{S_2^2} = \frac{0.216}{0.397} = 0.544.$$

对于给定 $\alpha = 0.05$, 查表 $F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0.025}(7, 6) = 5.70$, $F_{0.025}(6, 7) = 5.12$,

$$F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0.975}(7, 6) = \frac{1}{F_{0.025}(6, 7)} = \frac{1}{5.12} = 0.195.$$

因为 $F_{0.975}(7, 6) < F < F_{0.025}(7, 6)$, 故认为两个正态总体方差无显著差异.

例 7.3.6 设用两种不同方法冶炼某种金属材料, 分别抽样测定其杂质的含量(单位百分率)数据如下表. 试问新冶炼方法较原冶炼方法杂质含量的差异性是否有显著的降低.

原冶炼方法	22.3	24.2	26.1	26.4	26.9	22.8	25.7	23.0	27.2
杂质含量	30.2	24.5	29.5	25.1					
新冶炼方法	22.6	22.5	20.6	23.5	24.3	21.9	20.6		
杂质含量	23.2	23.4							

解 原方法与新方法的杂质含量分别记作 X 及 Y , 假定 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$. 于是提出如下的假设检验:

$$H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 > \sigma_2^2.$$

$$n_1 = 13, \bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i = \frac{1}{13} (22.3 + 24.2 + \dots + 25.1) = 25.68,$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 = \frac{1}{12} [(22.3 - 25.68)^2 + (24.2 - 25.68)^2 + \dots + (25.1 - 25.68)^2] = 6.178,$$

$$n_2 = 9, \bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i = \frac{1}{9} (22.6 + 22.5 + \dots + 23.4) = 22.51,$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2 = \frac{1}{8} [(22.6 - 22.51)^2 + (22.5 - 22.51)^2 + \dots + (23.4 - 22.51)^2] = 1.652.$$

$$F = \frac{S_1^2}{S_2^2} = \frac{6.178}{1.652} = 3.74.$$

给定 $\alpha = 0.05$, 查表临界值 $F_{\alpha}(n_1 - 1, n_2 - 1) = F_{0.05}(12, 8) = 3.28$.

因为 $F = 3.74 > 3.28 = F_{0.05}(12, 8)$, 所以拒绝 H_0 , 从而接受 H_1 . 即认为新冶炼方法的杂质含量的差异性比原冶炼方法降低是显著的.

例 7.3.7 比较甲、乙两种安眠药的治疗效果. 现独立观察 20 个病人, 其中 10 人服甲药, 另 10 人服乙药. 记录他们服药后睡眠延长时数, 算得两组平均睡眠延长时间和样本方差分别为 $\bar{x} = 1.80$, $s_1^2 = 3.52$; $\bar{y} = 0.75$, $s_2^2 = 2.88$. 问这两种药物的药效有无显著差异 ($\alpha = 0.1$)?

解 设 X 和 Y 分别为失眠者服用甲药和乙药后睡眠延长的时数, 并假定 X, Y 服从正态分布. 由于两个总体的方差未知, 利用 t 检验法比较两总体的均值时, 必须先检验它们的方差是否相等, 即先检验假设

$$H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2.$$

$$F = \frac{S_1^2}{S_2^2} = \frac{3.52}{2.88} = 1.22.$$

对于 $\alpha = 0.1$, $m_1 = n_2 = 10$, 查 F 分布表

$$F_{\frac{\alpha}{2}}(9, 9) = 3.18, F_{1-\frac{\alpha}{2}}(9, 9) = \frac{1}{F_{\frac{\alpha}{2}}(9, 9)} = \frac{1}{3.18} = 0.31.$$

由于 $F_{1-\frac{\alpha}{2}}(9,9)=0.31 < F=1.22 < 3.18 = F_{\frac{\alpha}{2}}(9,9)$.

故接受 H_0 , 即认为方差 $\sigma_1^2 = \sigma_2^2$.

再检验假设

$$H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2.$$

$$\begin{aligned} t &= \frac{\bar{x} - \bar{y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\bar{x} - \bar{y}}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \\ &= \frac{1.80 - 0.75}{\sqrt{10 \times (3.52 + 2.88)}} \sqrt{\frac{10 \times 10 \times 18}{20}} = 1.23. \end{aligned}$$

自由度 $n_1 + n_2 - 2 = 18$, 查 t 分布表 $t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) = 1.73$, 因为 $|t| = 1.23 < 1.73$.

故接受 H_0 , 即认为两种药物的疗效没有显著差异.

4. 成对数据的比较(t 检验法)

对成对数据差异性的检验常使用 t 检验法. 解题步骤为: ①提出假设; ②计算统计量 $T = \frac{\bar{d}}{S \sqrt{n}}$; ③确定临界值; ④做出检验判断.

例 7.3.8 9 名运动员在初进校时和接受训练一个星期后各进行一次体能测试, 测试评分为:

运动员 i	1	2	3	4	5	6	7	8	9
入学时 x_i	76	71	57	49	70	69	26	65	59
训练后 y_i	81	85	52	52	70	63	33	83	62
$d_i = x_i - y_i$	-5	-14	5	-3	0	6	-7	-18	-3

假设分数服从正态分布, 试在显著性水平 $\alpha = 0.05$ 下, 判断运动员的体能训练效果是否显著?

解 设 X, Y 分别表示入学时与训练后的分数, 显然测试结果 $x_i, y_i (i = 1, 2, \dots, 9)$ 不独立.

令 $d = X - Y$, 则 $d \sim N(\mu_d, \sigma^2)$. 据题意需检验假设:

$$H_0: \mu_d = 0, H_1: \mu_d < 0.$$

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \frac{1}{9} [-5 - 14 + \dots + (-3)] = -4.333,$$

$$\begin{aligned} s &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2} = \sqrt{\frac{1}{8} [(-5 + 4.333)^2 + (-14 + 4.333)^2 + \dots + (-3 + 4.333)^2]} \\ &= 7.937. \end{aligned}$$

$$t = \frac{\bar{d}}{S \sqrt{n}} = \frac{-4.333}{7.937 \sqrt{9}} = -1.638.$$

本题中 $n = 9, \alpha = 0.05$, 查表得临界值

$$-t_{\alpha}(n-1) = -t_{0.05}(8) = -1.8595,$$

因为 $t = -1.638 > -1.8595$, 所以接受 H_0 , 即可以认为体能训练效果不显著.

例 7.3.9 为了比较用来做鞋子后跟的两种材料质量,选取了 15 个男子(他们的生活条件各不相同),每人穿一双新鞋,其中一只以材料 A 做后跟,另一只以材料 B 做后跟,其厚度均为 10 mm.过了一个月再测量厚度,得到数据如下:

男子 i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
材料 A x_i	6.6	7.0	8.3	8.2	5.2	9.3	7.9	8.5	7.8	7.5	6.1	8.9	6.1	9.4	9.1
材料 B y_i	7.4	5.4	8.8	8.0	6.8	9.1	6.3	7.5	7.0	6.5	4.4	7.7	4.2	9.4	9.1

设 $d_i = X_i - Y_i (i=1, 2, \dots, 15)$ 是来自正态总体 $N(\mu_d, \sigma^2)$ 的样本; μ_d, σ^2 均未知.问是否可以认为以材料 A 制成的后跟比材料 B 的耐穿($\alpha=0.05$)?

解 本题属成对数据的比较,检验假设为

$$H_0: \mu_d \leq 0, H_1: \mu_d > 0.$$

$n=15$ $d_i = x_i - y_i (i=1, 2, \dots, 15)$ 计算得

$$\begin{array}{cccccccc} -0.8 & 1.6 & -0.5 & 0.2 & -1.6 & 0.2 & 1.6 & 1.0 \\ 0.8 & 1.0 & 1.7 & 1.2 & 1.9 & 0 & 0 & \end{array}$$

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \frac{1}{15} (-0.8 + 1.6 + \dots + 0) = 0.5533,$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2} = 1.0225,$$

$$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{0.5533}{1.0225/\sqrt{15}} = 2.0958.$$

对于 $\alpha=0.05$,查表得临界值 $t_{\alpha}(n-1) = t_{0.05}(14) = 1.7613$

因为 $t = 2.0958 > 1.7613 = t_{0.05}(14)$,故拒绝 H_0 .认为材料 A 的制品较材料 B 的耐穿.

同步练习 7.3

1. 设总体 $X \sim N(\mu_1, \sigma_1^2)$, 总体 $Y \sim N(\mu_2, \sigma_2^2)$, 其中 σ_1^2, σ_2^2 已知. 设 $x_1, x_2, \dots, x_{n_1}$ 是来自总体 X 的样本, $y_1, y_2, \dots, y_{n_2}$ 是来自总体 Y 的样本, 两样本独立, 则对于假设检验 $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$ 使用的统计量 u 为_____.

2. 两个正态总体 X 与 Y 的方差 σ_1^2 与 σ_2^2 未知, 则检验假设 $H_0: \mu_1 \leq \mu_2$, 要求 σ_1^2 与 σ_2^2 应该满足的条件是()

A. $\sigma_1^2 < \sigma_2^2$

B. $\sigma_1^2 > \sigma_2^2$

C. $\sigma_1^2 = \sigma_2^2$

D. σ_1^2, σ_2^2 可为任意正数

3. 为了化验甲、乙两种香烟中尼古丁的含量是否相同,现从甲、乙香烟中各抽取 6 份样品进行化验,测得尼古丁含量(单位:mg)为:

甲 23 24 27 26 21 23

乙 26 27 23 31 23 25

据经验知尼古丁含量服从正态分布,且甲种方差为 5,乙种方差为 7.问两种烟卷尼古丁含量是否有显著差异($\alpha=0.05$).

4. 为比较 A, B 两棉纺厂棉纱的断裂强度(单位: kg), 从两厂产品中各取一个样本, 测试其断裂强度得如下数据:

$$A: n_1 = 200, \bar{x} = 0.532;$$

$$B: n_2 = 100, \bar{y} = 0.576.$$

若两厂棉纱强度 X, Y 分别服从 $\sigma_1 = 0.218, \sigma_2 = 0.176$ 的正态分布, 试检验这两个厂棉纱强度的均值有无显著差异($\alpha = 0.1$).

5. 在一台自动车床上加工直径为 2.050 mm 的轴, 假设轴直径的分布是正态的. 为了考察车床的工作状况, 现每隔两小时随机抽取容量为 10 的样本, 所得数据列于下表. 试分析该台车床工作是否稳定($\alpha = 0.01$).

零件编号	1	2	3	4	5	6	7	8	9	10
样本 1	2.066	2.063	2.068	2.060	2.067	2.063	2.059	2.062	2.062	2.060
样本 2	2.063	2.060	2.057	2.056	2.059	2.058	2.062	2.059	2.059	2.057

6. 改进某种金属的热处理方法, 要检验抗拉强度(单位: kg/cm²)有无显著提高. 在改进前取 12 个试样, 测量并计算得 $\bar{x}_1 = 28.67, (n_1 - 1)s_1^2 = 66.64$, 在改进后又取 12 个试样, 测量并计算得 $\bar{x}_2 = 31.75, (n_2 - 1)s_2^2 = 112.25$. 假定热处理前与热处理后金属抗拉强度分别为正态总体, 且总体方差相等. 问热处理后抗拉强度有无显著提高($\alpha = 0.05$)?

7. 设随机变量 X_i 服从正态分布 $N(\mu_i, \sigma_i^2)$, μ_i, σ_i^2 均未知, $i = 1, 2$. X_1 与 X_2 相互独立. 现有 5 个 X_1 的观察值, 其样本方差 $s_1^2 = 7.505$, 有 4 个 X_2 的观察值, 其样本方差 $s_2^2 = 2.593$. 检验 X_1 与 X_2 的方差是否相等($\alpha = 0.01$).

8. 为比较不同季节出生的女婴体重的方差, 从某年 12 月和 6 月出生的女婴中分别随机抽取 6 名及 10 名, 测其体重如下(单位: g).

12 月 3 520 2 960 2 560 2 960 3 260 3 960

6 月 3 220 3 220 3 760 3 000 2 920 3 740 3 060 3 080 2 940 3 060

假定新生女婴体重服从正态分布, 问新生女婴体重的方差是否冬季的比夏季的小($\alpha = 0.05$)?

9. 随机地选了 8 个人, 分别测量了他们在早晨起床时和晚上就寝时的身高(cm), 得到以下的数据:

序号	1	2	3	4	5	6	7	8
早上 x_i	172	168	180	181	160	163	165	177
晚上 y_i	172	167	177	179	159	161	166	175
$d_i = x_i - y_i$	0	1	3	2	1	2	-1	2

$D_i = X_i - Y_i$ 来自正态总体. 问是否可以认为早晨的身高比晚上的身高要高($\alpha = 0.05$)?

同步练习 7.3 参考答案

$$1. \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad 2. C.$$

$$3. H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2.$$

$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$, $|u| = 1.27 < 1.96$ 接受 H_0 .

4. $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$.

$z_{\frac{\alpha}{2}} = z_{0.05} = 1.65$, $|u| = 1.88 > 1.65$ 拒绝 H_0 .

5. $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$.

$$t = \frac{\bar{x} - \bar{y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{2.063 - 2.059}{\sqrt{7.2 \times 10^{-6}} \times \sqrt{\frac{1}{10} + \frac{1}{10}}} = 3.3.$$

$$t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) = t_{0.005}(18) = 2.878.$$

$|t| = 3.3 > 2.878 = t_{0.005}(18)$ 故拒绝 H_0 .

6. $H_0: \mu_1 \geq \mu_2$, $H_1: \mu_1 < \mu_2$.

因为 $t = -2.647 < -1.734 = -t_{0.05}(22)$ 所以拒绝 H_0 .

7. $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$.

$$F = \frac{7.505}{2.593} = 2.89.$$

$$F_{0.05}(4, 3) = 9.12,$$

$$F_{0.975}(4, 3) = 0.15.$$

拒绝域为 $W = \{F \geq 9.12 \text{ 或 } F \leq 0.15\}$ 由于 $F \notin W$ 所以接受 H_0 .

8. 设 X, Y 分别表示冬夏两季新生女婴的体重, 依题意 X 与 Y 独立, 且 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$.

$$H_0: \sigma_1^2 \leq \sigma_2^2, H_1: \sigma_1^2 > \sigma_2^2.$$

$$F = \frac{s_1^2}{s_2^2} = \frac{505.667}{93.956} = 5.382.$$

查表得 $F_{\alpha}(n_1 - 1, n_2 - 1) = F_{0.05}(5, 9) = 3.84$.

因为 $F = 5.382 > 3.84 = F_{0.05}(5, 9)$ 所以拒绝 H_0 .

即可以认为新生女婴体重的方差冬季比夏季大.

9. $H_0: \mu_d \leq 0$, $H_1: \mu_d > 0$.

$$t = \frac{\bar{d}}{S_d / \sqrt{n}} = \frac{1.25}{1.2817 / \sqrt{8}} = 2.758.$$

$t_{0.05}(7) = 1.8946$. 由于 $t = 2.758 > 1.8946 = t_{0.05}(7)$ 所以拒绝 H_0 , 认为早晨的身高比晚上的身高要高.

7.4 总体分布函数的假设检验

7.4.1 知识要点

1. 总体分布函数的假设检验

在总体的分布为未知时, 根据样本 $X_1, X_2, \dots, X_n$ 来检验关于总体分布的假设.

H_0 : 总体 X 的分布函数为 $F(x)$,

H_1 : 总体 X 的分布函数不是 $F(x)$,

这种问题称为总体分布函数的假设检验,或称分布函数的拟合检验.

这里,若 X 为离散型随机变量,则待检假设为分布律 $P\{X=x_i\}=p_i(i=1,2,\dots)$;若 X 为连续型随机变量,则待检假设为概率密度 $p(x)$.

2. 皮尔逊(Pearson) χ^2 检验

χ^2 检验法的基本思想如下:将随机试验的可能结果的全体分为 k 个互不相容的事件 $A_1, A_2, \dots, A_k$,在 H_0 成立的条件下计算 $P(A_i)=p_i, i=1,2,\dots,k$.在 n 次试验中,事件出现的频率 $\frac{n_i}{n}$ 与 p_i 的差异 $\left| \frac{n_i}{n} - p_i \right|$ 在 H_0 为真且试验次数较多时应该比较小.因此,引入统计量

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^k \frac{n_i^2}{np_i} - n,$$

并有如下定理(皮尔逊 1900 年提出).

定理 若 n 充分大($n \geq 50$),则当 H_0 为真时,不论总体 X 服从何种分布,统计量 $\chi^2 = \sum_{i=1}^k (n_i - np_i)^2 / np_i$ 近似地服从自由度为 $k - r - 1$ 的 χ^2 分布.其中 r 是分布中未知参数的个数.

上述定理的结论适用于作为分布检验判断,这种检验法称为皮尔逊 χ^2 检验.

3. χ^2 分布拟合检验的步骤

(1)提出零假设 H_0 : X 服从分布 $F(x)$;

(2)在 H_0 为真的条件下,利用样本值估计(如极大似然估计) $F(x)$ 中的未知参数;

(3)将实数轴分成 k 个不相交的区间 $(a_0, a_1], (a_1, a_2], \dots, (a_{k-1}, a_k]$,其中 a_0 可取至 $-\infty$, a_k 可取至 $+\infty$,一般有 $5 \leq k \leq 16$;

(4)计算样本值频数 n_i ,即 n 个样本 $x_1, x_2, \dots, x_n$ 落入第 i 个区间 $(a_{i-1}, a_i]$ 中的个数($i=1, 2, \dots, k$);

(5)在 H_0 为真的条件下,计算 X 落入各区间的概率 $p_i = P\{a_{i-1} < X \leq a_i\} = F(a_i) - F(a_{i-1})$,进而得到理论频数 $np_i(i=1, 2, \dots, k)$;

(6)将 n_i, np_i 代入 $\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$ 求出其值;

(7)对于显著性水平 α ,查 χ^2 分布表得临界值 $\chi_\alpha^2(k - r - 1)$;

(8)下结论,若 $\chi^2 \geq \chi_\alpha^2(k - r - 1)$ 则拒绝 H_0 ,否则接受 H_0 .

7.4.2 典型题型解析

例 7.4.1 以下是某妇产科医院在 91 天中诞生的婴儿数:

一天中诞生 的婴儿数 x_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
有 x_i 个婴儿 诞生的天数	0	0	1	0	2	5	4	6	0	6	8	0	8	9	12	7	0	6	4	0	0	4	0	3	0	2	2	0	1	0	1

试检验假设 :一天中诞生的婴儿数 X 服从正态分布($\alpha=0.025$).

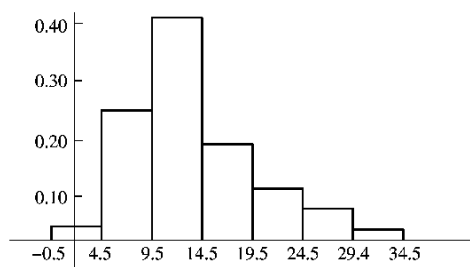
解 为粗略了解 X 的观察值的分布情况,先根据所给数据画出直方图.一天中诞生的婴儿数最小值为 0,最大值为 30,现取区间 $[-0.5, 34.5]$ 将它等分为 7 个小区间,小区间的长度为

$$\Delta = \frac{34.5 - (-0.5)}{7} = 5.$$

Δ 为组距,数出落在每个小区间内的数据的频数 n_i ,算出频率 $\frac{n_i}{n}$ ($n=91, i=1, 2, \dots, 7$),如下表

组号 i	组限	频数 n_i	频率 $\frac{n_i}{n}$	累积频率
1	$(-0.5, 4.5]$	3	0.033	0.033
2	$(4.5, 9.5]$	21	0.231	0.264
3	$(9.5, 14.5]$	37	0.407	0.671
4	$(14.5, 19.5]$	17	0.187	0.858
5	$(19.5, 24.5]$	7	0.077	0.935
6	$(24.5, 29.5]$	5	0.055	0.990
7	$(29.5, 34.5]$	1	0.010	1

自左至右依次在各小区间上作以 $\frac{n_1}{n} / \Delta$ 为高的小矩形,得到如下直方图:



H_0 : X 的概率密度为 $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbf{R}$.

用极大似然估计法估计未知参数 μ, σ^2 .

$$\hat{\mu} = \bar{x} = 13.27, \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = (5.85)^2.$$

此时待检假设 $H_0: X \sim N(13.27, (5.85)^2)$.

计算小区间上概率 $p_i = P\{a_{i-1} < X \leq a_i\} = \Phi\left(\frac{a_i - \hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{a_{i-1} - \hat{\mu}}{\hat{\sigma}}\right)$.

例如 $p_2 = P\{4.5 < X \leq 9.5\} = \Phi\left(\frac{9.5 - 13.27}{5.85}\right) - \Phi\left(\frac{4.5 - 13.27}{5.85}\right) = 0.1943$.

理论频数 $np_i = 91 \times 0.1943 = 17.6813$,

$$\frac{(n_i - np_i)^2}{np_i} = 0.6229.$$

其他情形类似计算,列表计算结果如下:

组号 i	n_i	p_i	np_i	$(n_i - np_i)^2 / np_i$
1	3	0.066 8	6.078 8	1.559 4
2	21	0.194 3	17.681 3	0.622 9
3	37	0.322 1	29.311 1	2.017
4	17	0.272 2	24.770 2	2.437 4
5	7	0.117 2	10.665 2	13.158 6
6	5	0.024 6	2.238 6	
1	1	0.002 8	0.254 8	
合计				6.638 6 ($=\chi^2$)

$\alpha = 0.025$ 临界值 $\chi^2_{\alpha}(k-r-1) = \chi^2_{0.025}(5-2-1) = 7.378 > 6.638 6 = \chi^2$, 所以接受 H_0 , 即认为一天中诞生的婴儿数 X 服从正态分布.

注 ①在用 χ^2 拟合检验实际问题时, 可先画出直方图, 这样可以大致了解数据的分布情况;

②若 np_i 小于等于 5 时, 要将 A_i (分组) 合并;

③未知参数只允许用极大似然估计法得到的估计值来代替.

例 7.4.2 考察某电话交换站一天中电话接错次数 X . 统计 267 天的记录, 各天电话接错次数的频数分布列成下表:

i (一天中电话接错次数)	n (天数)	np_i	i (一天中电话接错次数)	n (天数)	np_i
0~2	1 } ₆	2.05 } _{6.81}	10	35	30.63
3	5 } ₆	4.76 } _{6.81}	11	20	24.34
4	11	10.39	12	18	17.72
5	14	18.16	13	12	11.92
6	22	26.45	14	7	7.44
7	43	33.03	15	6 } ₈	4.33 } _{8.98}
8	31	36.09	≥ 16	2 } ₈	4.65 } _{8.98}
9	40	35.04			

问 X 是否符合泊松分布 ($\alpha = 0.05$).

解 先求泊松分布中的参数 λ , 用极大似然估计.

$$\hat{\lambda} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{267} (2 \times 1 + 3 \times 5 + \dots + 15 \times 6 + 16 \times 2) = 8.74.$$

然后计算理论频数 $np_i = 267 \times \frac{(8.74)^i}{i!} e^{-8.74}$,

计算结果列于上表. 且注意理论频数小于 5 的合并, 得到组数 $k = 13$. 计算得

$$\chi^2 = \sum_{i=1}^{13} \frac{(n_i - np_i)^2}{np_i} = 7.80.$$

$\alpha = 0.05$, 自由度 $k - r - 1 = 13 - 1 - 1 = 11$, 查表临界值 $\chi^2_{\alpha}(k-r-1) = \chi^2_{0.05}(11) = 19.675$.

因为 $\chi^2 = 7.80 < 19.675 = \chi^2_{0.05}(11)$ 故可认为 X 符合泊松分布.

同步练习 7.4

1. 下面列出了 84 位伊特拉斯坎(Etruscan)人男子的头颅的最大宽度(mm), 试检验这些

数据是否来自正态总体($\alpha=0.1$)?

141	148	132	138	154	142	150	146	155	158
150	140	147	148	144	150	149	145	149	158
143	141	144	144	126	140	144	142	141	140
145	135	147	146	141	136	140	146	142	137
148	154	137	139	143	140	131	143	141	149
148	135	148	152	143	144	141	143	147	146
150	132	142	142	143	153	149	146	149	138
142	149	142	137	134	144	146	147	140	142
140	137	152	145						

2. 自 1965 年 1 月 1 日至 1971 年 2 月 9 日共 2 231 天中,全世界记录到里氏震级 4 级和 4 级以上地震计 162 次,统计如下:

相继两次地震间隔天数 x	0—4	5—9	10—14	15—19	20—24	25—29	30—34	35—39	≥ 40
出现的频数	50	31	26	17	10	8	6	6	8

试检验相继两次地震间隔的天数 X 服从指数分布($\alpha=0.05$).

3. 检验一颗骰子的六个面是否匀称($\alpha=0.05$). 现在掷 120 次,结果如下:

点数	1	2	3	4	5	6
频数	21	28	19	24	16	12

同步练习 7.4 参考答案

1. 所给数据的最小值、最大值分别为 126、158,即所有数据落在区间 $[126, 158]$ 上,现取区间 $[124.5, 159.5]$,它能覆盖区间 $[126, 158]$ 将 $[124.5, 159.5]$ 等分为 7 个小区间,小区间长度 $\Delta = \frac{159.5 - 124.5}{7} = 5$,数落入每个小区间内的数据个数 n_i .

$$H_0: X \text{ 的概率密度 } p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty.$$

由极大似然估计参数 μ, σ^2 得

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 143, \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 6.0^2.$$

此时
$$p(x) = \frac{1}{\sqrt{2\pi} \times 6} e^{-\frac{(x-143.8)^2}{2 \times 6^2}}, -\infty < x < \infty.$$

$$p_i = P(A_i) = P\{a_{i-1} < X \leq a_i\} = \Phi\left(\frac{a_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_{i-1} - \mu}{\sigma}\right).$$

例如 $p_2 = P(A_2) = P\{129.5 < X \leq 134.5\}$

$$= \Phi\left(\frac{134.5 - 143.8}{6}\right) - \Phi\left(\frac{129.5 - 143.8}{6}\right)$$

$$= \Phi(-1.55) - \Phi(-2.38) = 0.0519.$$

计算结果列表如下:

A_i	n_i	p_i	np_i	n_i^2 / np_i
$A_1: x \leq 129.5$	1	0.008 7	0.73	5.09
$A_2: 129.5 < x \leq 134.5$	4	0.051 9	4.36	
$A_3: 134.5 < x \leq 139.5$	10	0.175 2	14.72	
$A_4: 139.5 < x \leq 144.5$	33	0.312 0	26.21	41.55
$A_5: 144.5 < x \leq 149.5$	24	0.281 1	23.61	24.40
$A_6: 149.5 < x \leq 154.5$	9	0.133 6	11.22	10.02
$A_7: 154.5 < x < \infty$	3	0.037 5	3.15	14.37
				$\sum = 87.67$

$$\chi^2 = \sum_{i=1}^k \frac{n_i^2}{np_i} - n = 87.67 - 84 = 3.67.$$

因为 $\chi_a^2(k-r-1) = \chi_{0.1}^2(5-2-1) = \chi_{0.1}^2(2) = 4.605 > 3.67 = \chi^2$,
故接受 H_0 , 即认为数据来自正态总体.

2. 按题意 $H_0: X$ 的概率密度为

$$p(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

未知参数 θ 用极大似然估计为 $\hat{\theta} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{2}{162} \frac{231}{162} = 13.77$.

在 H_0 下, X 可能取值的全体 Ω 为区间 $[0, +\infty)$, 将 $[0, +\infty)$ 分成 $k=9$ 个互不重叠的小区间: $A_1 = [0, 4.5], A_2 = [4.5, 9.5], \dots, A_9 = (39.5, \infty)$.

$$X \text{ 的密度函数 } p(x) = \begin{cases} \frac{1}{13.77} e^{-\frac{x}{13.77}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

$$\text{分布函数 } F(x) = \begin{cases} 1 - e^{-x/13.77}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

$$p_i = P(A_i) = P\{a_i < X \leq a_{i+1}\} = F(a_{i+1}) - F(a_i).$$

例如 $p_2 = P(A_2) = P\{4.5 < X \leq 9.5\} = F(9.5) - F(4.5) = 0.219 6$.

计算列表如下:

A_i	n_i	p_i	np_i	n_i^2 / np_i
$A_1: 0 \leq x \leq 4.5$	50	0.278 8	45.165 6	55.351 9
$A_2: 4.5 < x \leq 9.5$	31	0.219 6	35.575 2	27.013 2
$A_3: 9.5 < x \leq 14.5$	26	0.152 7	24.737 4	27.327 0
$A_4: 14.5 < x \leq 19.5$	17	0.106 2	17.204 4	16.798 0
$A_5: 19.5 < x \leq 24.5$	10	0.073 9	11.971 8	8.353 0
$A_6: 24.5 < x \leq 29.5$	8	0.051 4	8.326 8	7.686 0
$A_7: 29.5 < x \leq 34.5$	6	0.035 8	5.799 6	6.207 3
$A_8: 34.5 < x \leq 39.5$	6	0.024 8	4.017 6	14.826 9
$A_9: 39.5 < x < \infty$	8	0.056 8	9.201 6	13.219 2
				$\sum = 163.563 3$

$$\chi^2 = \sum_{i=1}^k \frac{n_i^2}{np_i} - n = 163.563 3 - 162 = 1.563 3.$$

因为 $\chi_a^2(k-r-1) = \chi_{0.05}^2(8-1-1) = \chi_{0.05}^2(6) = 12.592 > 1.563 3$.

所以接受 H_0 , 认为 X 服从指数分布.

3. 按题意,即检验假设.

	X	1	2	3	4	5	6
$H_0: X$ 的分布列为	p_k	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

其中 X 为骰子的点数.

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \frac{\left(21 - 120 \times \frac{1}{6}\right)^2}{120 \times \frac{1}{6}} + \frac{\left(28 - 120 \times \frac{1}{6}\right)^2}{120 \times \frac{1}{6}} + \dots + \frac{\left(12 - 120 \times \frac{1}{6}\right)^2}{120 \times \frac{1}{6}}$$

$$= 8.1.$$

查表得 $\chi_{\alpha}^2(k-r-1) = \chi_{0.05}^2(6-1) = \chi_{0.05}^2(5) = 11.07$.

因为 $\chi^2 = 8.1 < 11.07 = \chi_{0.05}^2(5)$, 所以接受 H_0 , 即可以认为骰子六个面是均匀的.

疑难解析

1. 在假设检验中,如何确定零假设 H_0 和备择假设 H_1 ?

答 在实际问题中,通常把需要着重考虑的假设视为零假设.具体体现为:①若问题是要决定新提出的方法是否比原方法好,则将原方法取为零假设 H_0 ,而将新方法取为备择假设 H_1 ;②若提出一个假设,检验的目的仅仅是为了判别这个假设是否成立,此时则直接取此假设为零假设 H_0 .

若从数学角度上看,原假设 H_0 与备择假设 H_1 的地位是平等的,但在实际问题中,如果提出的假设检验仅仅控制了犯第一类错误的概率,那么选用哪个假设作为零假设 H_0 ,要依具体问题的目的与要求而定.它取决于犯两类错误将会带来的后果,一般地可根据以下三个原则选择零假设 H_0 :①当目的是希望从样本观察值取得对某一诊断强有力的支持时,把这一诊断的否定作为零假设 H_0 ;②尽可能使后果严重的一类错误成为第一类错误;③把由过去资料所提供的诊断作为零假设,这样当检验后的最终结论为拒绝时,由于犯第一类错误的概率被控制而显得有说服力或危害较小.

2. 假设检验中,无论做出拒绝原假设或接受原假设,都有可能犯错误,这种说法对吗?

答 这种说法正确.即无论采取什么样的决策(取或舍)都可能是正确的,同时又都可能犯错误的.既然如此,假设检验的意义何在?

事实上,我们知道概率论本身就是研究随机现象的,因此它的结论无不带有随机性.小概率事件原理指出“小概率事件在一次试验中几乎不可能发生”,这里“几乎”就带有随机性.对原假设做出拒绝还是接受的判断,都是根据“小概率事件原理”,因此犯错误和不犯错误的可能性都是存在的,若二者的可能各占一半,那么“假设检验”确实没有任何价值.事实上,犯错误的概率是很小的,这样,“假设检验”才成为检验某种估计可靠程度的一种优良方法.

3. 如何合理地选取显著性水平 α ?

答 假设 H_0 成立,但由于样本的随机性,仍有可能做出拒绝 H_0 的结论,即犯第一类错误.显著性水平 α 的意义之一是给出了犯第一类错误的概率,即 H_0 成立, α 相应的临界值 λ_{α} ,

则统计量满足不等式 $|t| > \lambda_\alpha$ 的概率为 α . 另一方面, α 的选定又是对小概率事件小到何种程度的一种抉择. α 越小, 而事件发生了, 则拒绝 H_0 的可信程度越高, 所谓显著性即是指实际情况与 H_0 的判断之间存在显著差异. α 的选取通常取较小的值, 如 0.05, 0.01 等, 但在某些实际问题中, 如药品检验将不合格视为合格, 即犯第二类错误的后果更为严重时, 通常取 α 较大 (如 0.10), 使犯第二类错误的概率变小, 事实上犯两类错误的概率在样本容量固定时有此消彼长的关系.

4. 第一类错误和第二类错误的概率和为 1 吗?

答 不对. 若 α 为假设检验中犯第一类错误 (弃真) 的概率, β 为犯第二类错误 (取伪) 的概率. 若认为 $\alpha + \beta = 1$, 这是把发生第一类错误与发生第二类错误当成相互对立的事件. 实际上并非如此, 一般来说, α 应比较小, 而 β 应比 α 大, 但应该不至于大过 0.5, 即备择假设真时接受零假设的概率不大于拒绝零假设的概率. 还需指出, 当样本容量足够大时, 可以使 α, β 同时很小.

5. 检验的结果会随所给显著性水平的不同而改变吗?

答 会. 在同一个假设检验问题中, 对于零假设 H_0 , 当给定不同的显著性水平 α , 就对应着不同的临界值, 相应地有不同的拒绝域, 因而判断的结论不一定相同.

例如, 设 $X \sim N(\mu, 0.015^2)$, $X_1, X_2, \dots, X_9$ 为容量 $n=9$ 的样本, 若样本均值 $\bar{x} = 0.509$, 检验

$$H_0: \mu = 0.5, H_1: \mu \neq 0.5.$$

$$\text{令 } u = \frac{\bar{x} - 0.5}{0.015/\sqrt{9}}, \text{ 若 } H_0 \text{ 为真时, 则 } u \sim N(0, 1).$$

下面分别就显著性水平 (1) $\alpha_1 = 0.01$ (2) $\alpha_2 = 0.10$ 进行讨论.

(1) 当 $\alpha_1 = 0.01$ 时, 临界值 $z_{\alpha_1} = z_{0.01} = 2.58$ (查表),

$$u = \frac{\bar{x} - 0.5}{0.015/\sqrt{9}} = \frac{0.509 - 0.5}{0.015/\sqrt{9}} = 1.8.$$

因为 $|u| = 1.8 < z_{\alpha_1} = 2.58$, 故接受 H_0 , 即可认为 $\mu = 0.5$.

(2) 当 $\alpha_2 = 0.10$ 时, 临界值 $z_{\alpha_2} = z_{0.10} = 1.65$.

因为 $|u| = 1.8 > 1.65 = z_{\alpha_2}$, 所以拒绝 H_0 , 接受 H_1 , 即可认为 $\mu \neq 0.5$.

6. 假设检验与区间估计之间的异同是什么?

答 假设检验与区间估计对问题的提法虽不同, 但解决问题的途径是相同的. 现以正态总体 $N(\mu, \sigma_0^2)$ 的方差 σ_0^2 已知, 关于期望的区间估计与假设检验为例说明.

设 $X_1, X_2, \dots, X_n$ 是来自总体 $N(\mu, \sigma_0^2)$ 的样本, $x_1, x_2, \dots, x_n$ 是样本值, 则 μ 的置信度为 $1 - \alpha$ 的置信区间为 $\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \right)$, 即

$$P\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}\right) = 1 - \alpha.$$

$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$, 若 H_0 为真, $u = \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \sim N(0, 1)$, 对给定的显著性水平 α , 有

$$P\{|u| > z_{\frac{\alpha}{2}}\} = \alpha,$$

而 $P\{|u| < z_{\frac{\alpha}{2}}\} = 1 - \alpha.$

由此 H_0 的接受域为 $(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$, 也即对 μ 而言接受域为 $\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}\right)$, 就是说以 $1 - \alpha$ 的概率接受 H_0 . 而这个假设检验的接受域正是 μ 的置信度为 $1 - \alpha$ 的置信区间, 说明它们两者解决问题的途径是相同的, 参数的假设检验和参数的区间估计是从不同角度回答同一问题, 假设检验判断结论是否成立, 参数估计解决的是多少(或范围), 前者解决的是定性的, 后者解决是定“量”的.

例 设 $X \sim N(\mu, 1)$, μ 未知, $\alpha = 0.05$, $n = 16$, 且由样本算得 $\bar{x} = 5.20$, 于是 μ 的一个置信水平 $1 - \alpha = 0.95$ 的置信区间

$$\left(\bar{x} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{x} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}\right) = (5.20 - 0.49, 5.20 + 0.49) = (4.71, 5.69).$$

考虑检验问题 $H_0: \mu = 5.5$, $H_1: \mu \neq 5.5$, 由于 $5.5 \in (4.71, 5.69)$, 故接受 H_0 .

7. 为什么在两个总体情形, 检验假设 $H_0: \mu_1 = \mu_2$ 时, 若 σ_1^2, σ_2^2 未知, 要求 $\sigma_1^2 = \sigma_2^2 = \sigma^2$?

答 因为若 σ_1^2, σ_2^2 未知, 且 $\sigma_1^2 \neq \sigma_2^2$, 要检验假设 $H_0: \mu_1 = \mu_2$ 是不合理的, 是在不同条件下两总体均值的比较, 不能真实反映两总体的某数量指标的优劣. 只有令 $\sigma_1^2 = \sigma_2^2$, 才能在两总体的相同条件下比较某项数量指标的优劣, 做出的判断才有意义. 当 σ_1^2 与 σ_2^2 是否相等未知时, 要先用 F 检验法检验 σ_1^2 是否等于 σ_2^2 . 若相等, 再用 t 检验法检验两总体的期望 μ_1 和 μ_2 是否相等.

若 F 检验的结果是 $\sigma_1^2 \neq \sigma_2^2$, 这时要检验假设 $H_0: \mu_1 = \mu_2$, 有以下方法:

(1) 当 n_1, n_2 很大时, 选 $U = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim N(0, 1)$ 为统计量, 拒绝域为

$$|\bar{X} - \bar{Y}| \geq z_{\alpha/2} \sqrt{S_1^2/n_1 + S_2^2/n_2}.$$

(2) 当 $n_1 = n_2 = n$ 时, 令 $Z_i = X_i - Y_i$, $i = 1, 2, \dots, n$, 记 $E(Z_i) = \mu$, $D(Z_i) = \sigma^2$, 则 $Z_1, Z_2, \dots, Z_n$ 为 $N(\mu, \sigma^2)$ 的样本, $H_0: \mu = 0$. 检验统计量为

$$T = \bar{Z} \sqrt{n} / S_z^2, S_z^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2,$$

拒绝域为

$$|\bar{Z}| \geq t_{\alpha/2}(n-1) \frac{S_z^2}{\sqrt{n}}.$$

(3) 当 $\sigma_1^2 = k\sigma_2^2$ 时, 统计量为

$$T = \frac{|\bar{X} - \bar{Y}|}{S_w \sqrt{1/n_1 + 1/(kn_2)}} \sim t(n_1 + n_2 - 2), S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

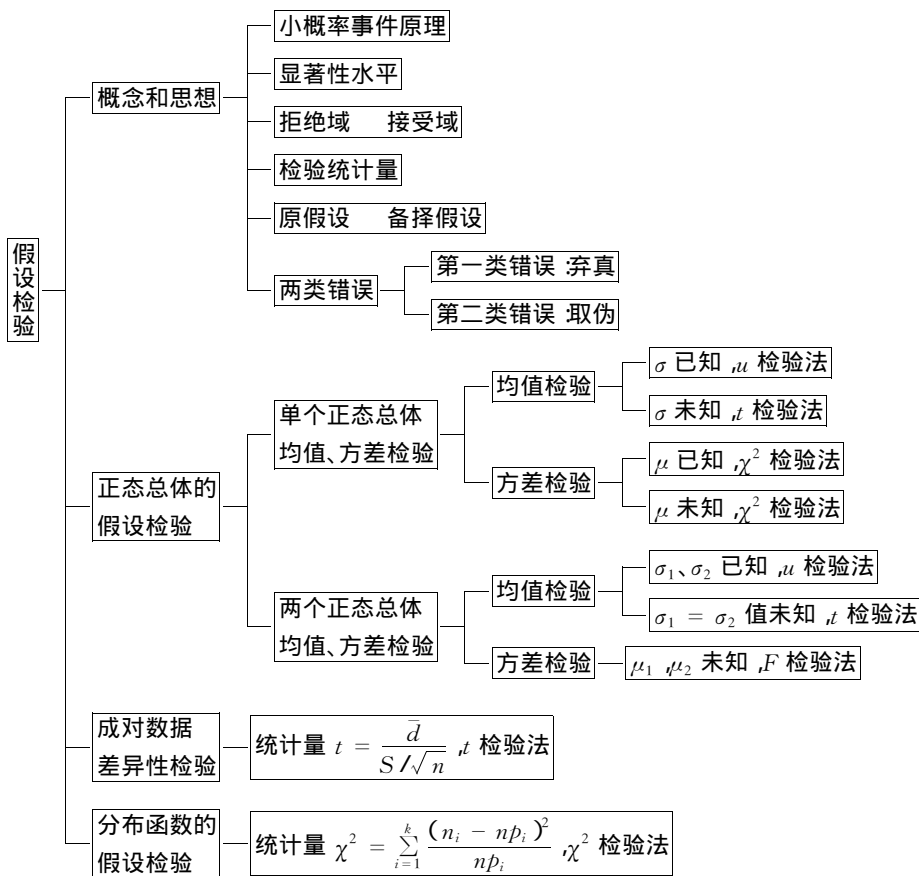
拒绝域为

$$|\bar{X} - \bar{Y}| \geq t_{\alpha/2}(n_1 + n_2 - 2) S_w \sqrt{1/n_1 + 1/(kn_2)}.$$

本章常考题型归纳

考点	常见题型	例题
假设检验基本概念	检验统计量的辨识	例 7.1.1
	原假设备择假设的确立	例 7.1.2
	拒绝域的形式	例 7.1.3
	两类错误(弃真、取伪)的认知	例 7.1.4 例 7.1.5
单个正态总体均值假设检验	已知 σ^2 检验 $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$	例 7.2.1
	已知 σ^2 检验 $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$	
	已知 σ^2 检验 $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$	例 7.2.2
	未知 σ^2 检验 $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$	例 7.2.6
	未知 σ^2 检验 $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$	
	未知 σ^2 检验 $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$	例 7.2.8
单个正态总体方差的检验	未知 μ 检验 $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$	例 7.2.9
	未知 μ 检验 $H_0: \sigma^2 \leq \sigma_0^2, H_1: \sigma^2 > \sigma_0^2$	例 7.2.11
	未知 μ 检验 $H_0: \sigma^2 \geq \sigma_0^2, H_1: \sigma^2 < \sigma_0^2$	
	已知 μ 检验 $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$	
	已知 μ 检验 $H_0: \sigma^2 \leq \sigma_0^2, H_1: \sigma^2 > \sigma_0^2$	例 7.2.12
	已知 μ 检验 $H_0: \sigma^2 \geq \sigma_0^2, H_1: \sigma^2 < \sigma_0^2$	
两个正态总体均值的检验	已知 σ_1^2, σ_2^2 检验 $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$	例 7.3.1
	已知 σ_1^2, σ_2^2 检验 $H_0: \mu_1 \leq \mu_2, H_1: \mu_1 > \mu_2$	例 7.3.2
	已知 σ_1^2, σ_2^2 检验 $H_0: \mu_1 \geq \mu_2, H_1: \mu_1 < \mu_2$	
	σ_1^2, σ_2^2 未知 $\sigma_1^2 = \sigma_2^2$ 检验 $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$	例 7.3.3
	σ_1^2, σ_2^2 未知 $\sigma_1^2 = \sigma_2^2$ 检验 $H_0: \mu_1 \leq \mu_2, H_1: \mu_1 > \mu_2$	
	σ_1^2, σ_2^2 未知 $\sigma_1^2 = \sigma_2^2$ 检验 $H_0: \mu_1 \geq \mu_2, H_1: \mu_1 < \mu_2$	例 7.3.4
两正态总体方差检验	μ_1, μ_2 未知检验 $H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$	例 7.3.7
	μ_1, μ_2 未知检验 $H_0: \sigma_1^2 \leq \sigma_2^2, H_1: \sigma_1^2 > \sigma_2^2$	例 7.3.6
	μ_1, μ_2 未知检验 $H_0: \sigma_1^2 \geq \sigma_2^2, H_1: \sigma_1^2 < \sigma_2^2$	
成对数据差异性检验	$H_0: \mu_1 - \mu_2 = 0 (\mu_d = 0), H_1: \mu_1 - \mu_2 \neq 0 (\mu_d \neq 0)$	
	$H_0: \mu_1 - \mu_2 \leq 0 (\mu_d \leq 0), H_1: \mu_1 - \mu_2 > 0 (\mu_d > 0)$	例 7.3.9
	$H_0: \mu_1 - \mu_2 \geq 0 (\mu_d \geq 0), H_1: \mu_1 - \mu_2 < 0 (\mu_d < 0)$	例 7.3.8
总体分布函数的假设检验	$H_0: \text{总体 } X \sim F(x), H_1: X \text{ 分布不是 } F(x)$	例 7.4.1 例 7.4.2

本章知识结构网络



检测题(一)

一、选择题

- 在假设检验中,显著性水平 α 的意义是().
 - 原假设 H_0 成立,经检验被拒绝的概率
 - 原假设 H_0 不成立,经检验被拒绝的概率
 - 原假设 H_0 成立,经检验不能拒绝的概率
 - 原假设 H_0 不成立,经检验不能拒绝的概率
- 在假设检验中,记 H_0 为原假设,则()为第一类错误.
 - H_0 为真,接受 H_0
 - H_0 不真,拒绝 H_0
 - H_0 为真,拒绝 H_0
 - H_0 不真,接受 H_0
- 设总体 $X \sim N(\mu, \sigma^2)$,其中 μ 未知,取样本 $x_1, x_2, \dots, x_n$,记 $\bar{x}, S^2$ 为样本均值与样本方差,对假设检验 $H_0: \sigma \geq 2, H_1: \sigma < 2$ 应取检验统计量 $\chi^2 = ()$.
 - $\frac{(n-1)S^2}{8}$
 - $\frac{(n-1)S^2}{2}$
 - $\frac{(n-1)S^2}{4}$
 - $\frac{(n-1)S^2}{6}$

4. 设 $\bar{X}$ 和 S^2 是来自正态总体 $N(\mu, \sigma^2)$ 的样本均值和样本方差, 样本容量为 n , $|\bar{x} - \mu_0| > t_{0.05}(n-1) \frac{S}{\sqrt{n}}$ 为().

A. $H_0: \mu = \mu_0$ 的拒绝域

B. $H_0: \mu = \mu_0$ 的接受域

C. μ 的一个置信区间

D. σ^2 的一个置信区间

5. 对正态总体 $N(\mu, \sigma^2)$ (σ^2 未知) 的假设检验问题: $H_0: \mu \leq 1, H_1: \mu > 1$, 若取显著性水平 $\alpha = 0.05$, 则其拒绝域为().

A. $|\bar{x} - 1| > z_{0.05}$

B. $\bar{x} > 1 + t_{0.05}(n-1) \frac{S}{\sqrt{n}}$

C. $|\bar{x} - 1| > t_{0.05} \frac{S}{\sqrt{n}}$

D. $\bar{x} < 1 - t_{0.05}(n-1) \frac{S}{\sqrt{n}}$

6. 设总体 $X \sim N(\mu_1, \sigma_1^2)$, $X_1, X_2, \dots, X_{n_1}$ 是来自 X 的简单随机样本, 总体 $Y \sim N(\mu_2, \sigma_2^2)$, $Y_1, Y_2, \dots, Y_{n_2}$ 是来自 Y 的简单随机样本, 且 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 均为未知参数, 两样本相互独立, 令

$$F_1 = \frac{(n_2 - 1) \sum_{i=1}^{n_1} (X_i - \bar{X})^2}{(n_1 - 1) \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}, F_2 = \frac{n_2 \sum_{i=1}^{n_1} (X_i - \mu_1)^2}{n_1 \sum_{i=1}^{n_2} (Y_i - \mu_2)^2},$$

则检验假设(给定显著性水平 α) $H_0: \sigma_1^2 \geq \sigma_2^2, H_1: \sigma_1^2 < \sigma_2^2$ 的拒绝域为().

A. $W = \{F_1 \leq F_{1-\alpha}(n_1-1, n_2-1)\}$

B. $W = \{F_1 \leq F_{\alpha}(n_1, n_2)\}$

C. $W = \{F_2 < F_{1-\alpha}(n_1, n_2)\}$

D. $W = \{F_2 < F_{\alpha}(n_1-1, n_2-1)\}$

7. 在统计假设中, 下面结论正确的是().

A. 单边检验的定义为“若提出原假设 $H_0: \mu = \mu_0$, 备择假设为 $\mu < \mu_0$, 则称该检验为单边检验”

B. 在假设检验中, 单边假设、双边假设是根据需要人为随意设定的, 与被检验的参数本身无关

C. 单边检验的定义为“若拒绝域 R_{α} 为 $\{-\infty < x < \lambda\}$ 称为单边检验

D. 上述说法都不对

8. 某工厂所生产的某种细纱支数服从正态分布 $N(\mu_0, \sigma_0^2)$, μ_0, σ_0^2 为已知, 现从某日生产的一批产品中, 随机抽 16 缕进行支数测量, 求得子样均值及方差为 A, B , 要检验纱的均匀度是否变劣, 则提出假设().

A. $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$

B. $H_0: \mu = \mu_0, H_1: \mu > \mu_0$

C. $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 > \sigma_0^2$

D. $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$

二、填空题

1. 设总体 $X \sim N(\mu, \sigma^2)$ (μ, σ^2 均未知), $X_1, X_2, \dots, X_n$ 是来自该总体的简单随机样本, 记 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, Q^2 = \sum_{i=1}^n (X_i - \bar{X})^2$, 则检验假设 $H_0: \mu = 0$ 时, 构造统计量为_____, 它服从_____分布.

2. 设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 其中 μ_1, μ_2 已知, $x_1, x_2, \dots, x_{n_1}$ 与 $y_1, y_2, \dots, y_{n_2}$ 分别是来自 X 与 Y 的样本, 两样本独立, 对假设检验 $H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$, 其检验统计量 $F = \underline{\hspace{2cm}}$, 其拒绝域 $W = \underline{\hspace{2cm}}$.

3. 从已知标准差 $\sigma = 5.2$ 的正态总体中, 抽取容量为 16 的样本, 算得样本均值 $\bar{x} = 27.56$, 在显著水平 $\alpha = 0.05$ 之下检验假设 $H_0: \mu = 26, H_1: \mu \neq 26$, 检验结果是 $\underline{\hspace{2cm}}$.

4. 假设检验的统计思想是概率很小的事件在一次试验中可以认为基本上是不会发生的, 该原理称为 $\underline{\hspace{2cm}}$.

三、解答题

1. 假设按某种工艺生产的金属纤维的长度(单位 mm) X 服从正态分布 $N(5.2, 0.16)$, 现在随机抽出 15 根纤维, 测得它们的平均长度 $\bar{x} = 5.4$, 如果估计方差没有变化, 可否认为现在生产的纤维平均长度仍为 5.2 mm ($\alpha = 0.05$)?

2. 设随机变量 X_i 服从正态分布 $N(\mu_i, \sigma_i^2)$, μ_i, σ_i^2 均未知, $i = 1, 2$. X_1 与 X_2 相互独立. 现有 5 个 X_1 的观察值, 其样本方差 $s_1^2 = 7.505$, 有 4 个 X_2 的观察值, 其样本方差 $s_2^2 = 2.593$, 且又知两个样本均值分别为 $\bar{x}_1 = 19, \bar{x}_2 = 18$, 如果 $\sigma_1^2 = \sigma_2^2$, 检验 μ_1 与 μ_2 是否相等 ($\alpha = 0.1$).

3. 下面列出的是某工厂随机选取的 20 只部件的装配时间(min):

9.8 10.4 10.6 9.6 9.7 9.9 10.9 11.1 9.6
10.2 10.3 9.6 9.9 11.2 10.6 9.8 10.5 10.1
10.5 9.7

设装配时间总体服从正态分布 $N(\mu, \sigma^2)$, μ, σ^2 均未知. 是否可以认为装配时间的均值 μ 显著大于 10 (取 $\alpha = 0.05$)?

4. 设某厂生产的缆绳, 其抗拉强度 $X \sim N(\mu, 82^2)$, 现在从改进工艺后生产的一批缆绳中随机抽取 10 根, 测量其抗拉强度, 算得样本均值 $\bar{x} = 10653$, 方差 $s^2 = 6992$, 当显著水平 $\alpha = 0.05$ 时, 能否据此样本认为新工艺生产的缆绳抗拉强度的方差有显著变化.

5. 某种导线, 要求其电阻的标准差不得超过 0.005 Ω . 今在生产的一批导线中取样品 9 根, 测得 $s = 0.007 \Omega$, 设总体为正态分布. 问在水平 $\alpha = 0.05$ 下能否认为这批导线的标准差显著地偏大?

6. 有两台机器生产金属部件, 现分别在两台机器所生产的部件中各取一容量 $n_1 = 60, n_2 = 40$ 的样本, 测得部件重量的样本方差分别为 $s_1^2 = 15.46, s_2^2 = 9.66$. 设两样本相互独立, 两总体分别服从 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ 分布. 试在水平 $\alpha = 0.05$ 下检验假设 $H_0: \sigma_1^2 \leq \sigma_2^2, H_1: \sigma_1^2 > \sigma_2^2$.

7. 下表分别给出两个文学家马克·吐温(Mark Twain)的 8 篇小品文以及斯诺特格拉斯(Snodgrass)的 10 篇小品文中内三个字母组成的词的比例.

马克·吐温	0.225	0.262	0.217	0.240	0.230	0.229	0.235	0.217		
斯诺特格拉斯	0.209	0.205	0.196	0.210	0.202	0.207	0.224	0.223	0.220	0.201

设两组数据分别来自正态总体, 且两总体方差相等. 两样本相互独立. 问两个作家所写的小品文中含由 3 个字母组成的词的比例是否有显著的差异 (取 $\alpha = 0.05$)?

8. 在一批灯泡中抽取 300 只做寿命试验, 其结果如下:

寿命 t (小时)	$t < 100$	$100 \leq t < 200$	$200 \leq t < 300$	$t \geq 300$
灯泡数	121	78	43	58

取 $\alpha = 0.05$, 试检验假设 H_0 : 寿命服从指数分布

$$p(x) = \begin{cases} 0.005e^{-0.005x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

检测题(一)参考答案

一、1. A 2. C 3. C 4. A 5. B 6. A 7. B 8. C

二、1. $\frac{\bar{X}}{Q} \sqrt{n(n-1)}$, $t(n-1)$

2. $\frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2 / \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$ $\{F < F_{1-\frac{\alpha}{2}}(n_1, n_2) \text{ 或 } F > F_{\frac{\alpha}{2}}(n_1, n_2)\}$

3. 接受 H_0 4. 小概率事件原理

三、1. $H_0: \mu = 5.2$, $H_1: \mu \neq 5.2$.

$$u = 1.94,$$

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96.$$

因为 $|u| = 1.94 < 1.96 = z_{\frac{\alpha}{2}}$, 接受 H_0 .

2. $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$,

$$S_w^2 = 5.4,$$

$$t = 0.64.$$

$$t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) = t_{0.05}(7) = 1.895.$$

因为 $|t| = 0.64 < 1.895 = t_{0.05}(7)$, 所以接受 H_0 .

3. 取 H_0 为维持原状, 即需检验假设为

$$H_0: \mu \leq 10, H_1: \mu > 10,$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{10.2 - 10}{0.5099/\sqrt{20}} = 1.754.$$

临界值 $t_{\alpha}(n-1) = t_{0.05}(19) = 1.729$. 因为 $t = 1.754 > 1.729 = t_{0.05}(19)$, 拒绝 H_0 .

4. $H_0: \sigma^2 = \sigma_0^2 = 82^2$, $H_1: \sigma^2 \neq 82^2$.

$$\chi^2 = 9.36.$$

$$\chi_{1-\frac{\alpha}{2}}^2(n-1) = \chi_{0.975}^2(9) = 2.7, \chi_{\frac{\alpha}{2}}^2(n-1) = \chi_{0.025}^2(9) = 19.02.$$

拒绝域为 $W = \{\chi^2 \leq 2.7 \text{ 或 } \chi^2 \geq 19.02\}$. 由于 $\chi^2 = 9.36 \notin W$, 故接受 H_0 .

5. $H_0: \sigma = \sigma_0 = 0.005$, $H_1: \sigma > \sigma_0$.

$$\chi^2 = 15.68,$$

$$\chi_{\alpha}^2(n-1) = \chi_{0.05}^2(8) = 15.507.$$

$$\chi^2 = 15.68 > 15.507 = \chi_{0.05}^2(8) \text{ 所以拒绝 } H_0.$$

6. $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 > \sigma_2^2$.

$$F = 1.60 ,$$

临界值 $F_{\alpha}(n_1 - 1, n_2 - 1) = F_{0.05}(59, 39) = 1.64$,

因为 $F = 1.60 < 1.64 = F_{0.05}(59, 39)$, 所以接受 H_0 .

7. 设马克·吐温小品文服从 $X \sim N(\mu_1, \sigma_1^2)$, 斯诺特格拉斯小品文 $Y \sim N(\mu_2, \sigma_2^2)$.

这里 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 均未知. 先检验

$$H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2 .$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 = 0.0146^2 ,$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2 = 0.0097^2 ,$$

$$F = \frac{s_1^2}{s_2^2} = \frac{0.0146^2}{0.0097^2} = 2.2655 .$$

$$F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0.975}(79) = \frac{1}{F_{0.025}(9, 7)} = \frac{1}{4.82} = 0.207 ,$$

$$F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0.025}(79) = 4.20 .$$

即有 $F_{0.975}(79) < F = 2.2655 < F_{0.025}(79)$. 所以接受 H_0 , 认为两总体方差是相同的.

在上述基础上检验 $H'_0: \mu_1 = \mu_2, H'_1: \mu_1 \neq \mu_2$.

$$t = \frac{\bar{x} - \bar{y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.2319 - 0.2097}{0.012 \sqrt{\frac{1}{8} + \frac{1}{10}}} = 3.9 ,$$

$$t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) = t_{0.025}(16) = 2.1199 .$$

$$\text{由于 } |t| = 3.9 > 2.1199 = t_{0.025}(16) ,$$

所以拒绝 H'_0 .

8. 若 H_0 为真, X 的分布函数为

$$F(x) = \begin{cases} 1 - e^{-0.005x} , & x \geq 0 , \\ 0 , & x < 0 . \end{cases}$$

$$p_1 = P(A_1) = P\{t < 100\} = F(100) = 0.39347 ,$$

$$p_2 = P(A_2) = P\{100 \leq t < 200\} = F(200) - F(100) = 0.23865 ,$$

$$\text{类似 } p_3 = 0.14475, p_4 = 0.22313 .$$

计算结果如下表:

A_i	n_i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2 / np_i$
$A_1: t < 100$	121	0.39347	118.041	2.959	0.0742
$A_2: 100 \leq t < 200$	78	0.23865	71.595	6.405	0.5729
$A_3: 200 \leq t < 300$	43	0.14475	43.425	-0.425	0.0042
$A_4: t \geq 300$	58	0.22313	66.939	-8.939	1.3778
					$\sum = 2.0291$

$$\chi_{\alpha}^2(k - r - 1) = \chi_{0.05}^2(4 - 0 - 1) = \chi_{0.05}^2(3) = 7.815 .$$

$$\text{因为 } \chi^2 = 2.0291 < 7.815 = \chi_{\alpha}^2(k - r - 1) ,$$

所以接受 H_0 .

检测题(二)

一、选择题

1. 设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, $\bar{X}$ 是样本均值, 记 $S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, $S_2^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, $S_3^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2$, $S_4^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$, μ, σ^2 均未知, 若提出检验假设 $H_0: \mu = \mu_0$, 则选用统计量().

A. $T = \frac{\bar{X} - \mu_0}{\frac{S_1}{\sqrt{n-1}}}$

B. $T = \frac{\bar{X} - \mu_0}{\frac{S_2}{\sqrt{n-1}}}$

C. $T = \frac{\bar{X} - \mu_0}{\frac{S_3}{\sqrt{n}}}$

D. $T = \frac{\bar{X} - \mu_0}{\frac{S_4}{\sqrt{n}}}$

2. 测定某溶液中的水分, 由它的 9 个测定值, 算出样本均值及样本标准差分别为 $\bar{x} = 0.452\%$, $s = 0.037\%$, 总体服从正态分布, 下面提出的检验假设被接受的是()

A. 在 $\alpha = 0.05$ 之下, 检验假设 $H_0: \mu = 0.5\%$

B. 在 $\alpha = 0.05$ 之下, 检验假设 $H_0: \sigma = 0.03\%$

C. 在 $\alpha = 0.25$ 之下, 检验假设 $H_0: \mu = 0.5\%$

D. 在 $\alpha = 0.25$ 之下, 检验假设 $H_0: \sigma = 0.03\%$

3. 对显著水平 α 检验结果而言, 犯第一类(去真)错误的概率 $P(\text{拒绝 } H_0 | H_0 \text{ 为真}) =$ ()

A. 不是 α

B. $1 - \alpha$

C. 大于 α

D. 小于或等于 α

4. 设 $X_1, X_2, \dots, X_n$ 是来自总体 $N(10, \sigma^2)$ 的样本, 针对 $H_0: \sigma^2 \leq 100$, $H_1: \sigma^2 > 100$, $\alpha = 0.05$. 关于此检验问题, 下列不正确的是()

A. 若设 W 为拒绝域, 则 $P\{(X_1, X_2, \dots, X_n) \in W | \sigma^2 \leq 100\} = 0.05$ 恒成立

B. 检验统计量取作 $\frac{(n-1)S^2}{100}$

C. 拒绝域可取为 $\left\{ \frac{\sum_{i=1}^n (X_i - 10)^2}{100} < k \right\}$ 的形式

D. 在 H_0 成立时, $\frac{\sum_{i=1}^n (X_i - 10)^2}{100}$ 服从 $\chi^2(n)$ 分布

二、填空题

1. 设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, 其中参数 μ 和 σ^2 未知, 若计算出

$$\bar{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 14.75, \sum_{i=1}^{16} (x_i - \bar{x})^2 = 53.5,$$

则假设 $H_0: \mu = 15$ 的 t 检验选用的 t 统计量的值 $t =$ _____.

2. 设总体 $X \sim N(\mu, \sigma^2)$ 检验假设 $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2, x_1, x_2, \dots, x_{11}$ 是一组样本观察值, 显著性水平 $\alpha = 0.05$, 则拒绝域是_____.

3. 某产品以往废品率不高于 5%, 今抽取一样本, 以检验这批产品废品率是否高于 5% (显著性水平 α). 此问题的假设为 H_0 : _____, 犯第一类错误的概率为_____.

三、解答题

1. 在某年级学生中抽测 9 名跳远成绩, 得样本均值 $\bar{x} = 4.38$ m. 假设跳远成绩 X 服从正态分布, 且 $\sigma = 0.3$, 问是否可以认为该年级学生跳远平均成绩为 $\mu = 4.40$ m ($\alpha = 0.10$).

2. 已知钢筋强度 X 服从正态分布, 且 $E(X) = 52$, 今改变炼钢的配方, 用新配方炼了 7 炉钢, 从这 7 炉钢生产的钢筋中分别各取一根, 测得样本均值为 52.14, 样本方差为 2.70^2 , 问用新法炼钢生产的钢筋强度的均值是否明显提高 ($\alpha = 0.05$)?

3. 某厂生产的某种型号的电池, 其寿命长期以来服从方差 $\sigma^2 = 5000$ (单位 h^2) 的正态分布, 现有一批这种电池, 从它的生产情况来看, 寿命的波动性有所改变. 现随机取 26 只电池, 测出其寿命的样本方差 $s^2 = 9200$ (单位 h^2). 根据这一数据推断这批电池的寿命的波动性较以往的有无显著的变化 ($\alpha = 0.02$).

4. 某校从经常参加体育锻炼的男生中随机地选出 50 名, 测得平均身高 174.34 cm; 从不经常参加体育锻炼的男生中随机地选 50 名, 测得平均身高 172.42 cm. 统计资料表明两种男生的身高都服从正态分布, 其标准差分别为 5.35 cm, 6.11 cm, 问该校经常参加锻炼的男生是否比不常参加体育锻炼的男生平均身高要高些 ($\alpha = 0.05$)?

5. 在 70 年代后期人们发现, 酿造啤酒时, 在麦芽干燥过程中形成致癌物质亚硝基二甲胺 (NDMA). 到了 80 年代初期开发了一种新的麦芽干燥过程. 下面给出分别在新老两种过程中形成的 NDMA 含量 (以 10 亿份中的份数计)

老过程	6	4	5	5	6	5	5	6	4	6	7	4
新过程	2	1	2	2	1	0	3	2	1	0	1	3

设两样本分别来自正态总体, 且两总体的方差相等. 两样本独立. 分别以 μ_1, μ_2 记对应老、新过程的总体的均值, 试检验假设 ($\alpha = 0.05$) $H_0: \mu_1 - \mu_2 = 2, H_1: \mu_1 - \mu_2 > 2$.

6. 甲、乙相邻两地段各取了 50 块和 52 块岩心进行磁化率测定, 算出 $\frac{1}{50} \sum_{i=1}^{50} (x_i - \bar{x})^2 = 0.139, \frac{1}{52} \sum_{i=1}^{52} (y_i - \bar{y})^2 = 0.0053$, 试问甲、乙两段的标准差是否有显著差异 ($\alpha = 5\%$).

7. 为了试验两种不同的谷物的种子的优劣, 选取了 10 块土质不同的土地, 并将每块土地分为面积相同的两部分, 分别种植这两种种子, 设在每块土地的两部分人工管理等条件完全一样, 下面给出各块土地上的产量.

土地	1	2	3	4	5	6	7	8	9	10
种子 A (x_i)	23	35	29	42	39	29	37	34	35	28
种子 B (y_i)	26	39	35	40	38	24	36	27	41	27

设 $d_i = x_i - y_i$ ($i = 1, 2, \dots, 10$) 来自正态总体, 问以这两种种子种植的谷物的产量是否有显著差异 ($\alpha = 0.05$)?

8. 袋中装有 8 只球, 其中红球数未知, 在其中任取 3 只, 记录红球的只数 X , 然后放回, 再任取 3 只, 记录红球的只数, 然后放回, 如此重复进行了 112 次, 其结果如下:

X	0	1	2	3
次数	1	31	55	25

试取 $\alpha = 0.05$ 检验假设 $H_0: X$ 服从超几何分布

$$P\{X=k\} = \frac{C_5^k C_3^{3-k}}{C_8^3}, k=0, 1, 2, 3,$$

即检验假设 H_0 红球的只数为 5.

检测题(二)参考答案

一、1. B 2. B 3. D 4. C

二、1. -0.53 2. $(-\infty, 3.35] \cup [20.5, +\infty)$

3. $p \leq 5\%$ (p 为废品率) α

三、1. $H_0: \mu = 4.40$, $H_1: \mu \neq 4.40$.

$$u = 0.2.$$

$$z_{\frac{\alpha}{2}} = z_{0.05} = 1.64.$$

因为 $|u| = 0.2 < 1.64 = z_{0.05}$, 所以接受 H_0 .

2. $H_0: \mu = 52$, $H_1: \mu > 52$.

$$t_{\alpha}(n-1) = t_{0.05}(6) = 1.94.$$

拒绝域 $W = \{t > 1.94\}$.

因为 $t = 0.14 < 1.94 = t_{0.05}(6)$, 故接受 H_0 .

3. $H_0: \sigma^2 = \sigma_0^2 = 5000$, $H_1: \sigma^2 \neq \sigma_0^2 = 5000$.

$$\chi_{\frac{\alpha}{2}}^2(n-1) = \chi_{0.01}^2(25) = 44.314.$$

$$\chi_{1-\frac{\alpha}{2}}^2(n-1) = \chi_{0.99}^2(25) = 11.524.$$

$$\chi^2 = 46,$$

由于 $\chi^2 = 46 > \chi_{0.01}^2(25) = 44.314$, 故拒绝 H_0 .

4. X, Y 分别表示常锻炼和不常锻炼男生的身高, 由题设

$$X \sim N(\mu_1, 5.35^2), Y \sim N(\mu_2, 6.11^2)$$

$$H_0: \mu_1 \leq \mu_2, H_1: \mu_1 > \mu_2.$$

$$u = 1.67.$$

$$z_{\alpha} = 1.64.$$

由于 $u = 1.67 > 1.64 = z_{0.05}$, 所以拒绝 H_0 .

$$5. t = \frac{\bar{x} - \bar{y} - 2}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{5.25 - 1.5 - 2}{\sqrt{\frac{11 \times 0.93182 + 11 \times 1}{22}} \sqrt{\frac{1}{12} + \frac{1}{12}}} = 4.3616.$$

查表临界值 $t_{\alpha}(n_1 + n_2 - 2) = t_{0.05}(22) = 1.7207$.

由 $t = 4.3616 > 1.7207 = t_{0.05}(22)$, 所以拒绝 H_0 , 即认为 $\mu_1 - \mu_2 > 2$.

6. $H_0: \sigma_1 = \sigma_2$, $H_1: \sigma_1 \neq \sigma_2$

$$F = \frac{s_1^2}{s_2^2} = \frac{0.0142}{0.0054} = 2.63.$$

$$F_{\frac{\alpha}{2}}(50-1, 52-1) = F_{0.025}(49, 51) = 1.76,$$

$$F_{1-\frac{\alpha}{2}}(50-1, 52-1) = F_{0.975}(49, 51) = \frac{1}{F_{0.025}(51, 49)} = 1.59.$$

由 $F = 2.63 > 1.76 = F_{\frac{\alpha}{2}}$, 故拒绝 H_0 .

7. 本题是成对数据的检验

$$H_0: \mu_d = 0, H_1: \mu_d \neq 0.$$

作差 $d_i = x_i - y_i$ 得

土地	1	2	3	4	5	6	7	8	9	10
d_i	-3	-4	-6	2	1	5	1	7	-6	1

检验统计量 $t = \frac{\bar{d}}{S/\sqrt{n}} \sim t(n-1)$ (在 H_0 成立时).

$$t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(9) = 2.2622.$$

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = -0.2, S = 4.44222,$$

$$|t| = \left| \frac{-0.2}{4.44222/\sqrt{10}} \right| = 0.1424.$$

由于 $|t| = 0.1424 < 2.2622 = t_{0.025}(9)$, 所以接受 H_0 .

8. 根据题意可知为检验假设

$$H_0: P\{X=k\} = \frac{C_5^k C_3^{3-k}}{C_8^3}, k=0, 1, 2, 3.$$

将试验可能结果的全体分成 $k=4$ 个两两互不相容事件 A_0, A_1, A_2, A_3 ,

$$p_i = P\{X=i\} = \frac{C_5^i C_3^{3-i}}{C_8^3},$$

列表计算如下:

A_i	n_i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2 / np_i$
A_0	1	0.01786	2.0003	0	0
A_1	31	0.26786	30		
A_2	55	0.53714	60	-5	0.4167
A_3	25	0.1786	20	5	1.25
					$\sum = 1.667$

因为 $\chi_a^2(k-r-1) = \chi_{0.05}^2(3-1) = \chi_{0.05}^2(2) = 5.991 > 1.667 = \chi^2$,

所以接受 H_0 .

第8章 方差分析

8.1 单因素方差分析

8.1.1 知识要点

1. 何谓方差分析

方差分析是分析试验(或观测)数据的一种方法,它所解决的基本问题是通过数据分析,弄清与研究对象有关的多个因素,以及各个因素之间交互作用对该对象的影响.它所研究的对象都假定服从正态分布.

2. 单因素方差分析数学模型

设因素 A 有 r 个水平 $A_1, A_2, \dots, A_r$, 在水平 $A_j (j=1, 2, \dots, r)$ 下进行 $n_j (n_j \geq 2)$ 次独立试验, 试验记录如下表:

水平 \ 样本	A_1	A_2	...	A_r
样本 1	X_{11}	X_{12}	...	X_{1r}
样本 2	X_{21}	X_{22}	...	X_{2r}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
样本 n_j	$X_{n_1 1}$	$X_{n_2 2}$	...	$X_{n_r r}$

其中 X_{ij} 表示第 j 水平 A_j 进行第 i 次试验的结果.

假定 $X_{ij} \sim N(\mu_j, \sigma^2)$ 其中 μ_j 与 σ^2 未知, 在水平 A_j 下样本相互独立. 则有

$$X_{ij} - \mu_j \sim N(0, \sigma^2),$$

从而 $\epsilon_{ij} = X_{ij} - \mu_j$ 为随机误差. 于是

$$\begin{cases} X_{ij} = \mu_j + \epsilon_{ij} \quad i=1, 2, \dots, n_j, \quad j=1, 2, \dots, r, \\ \epsilon_{ij} \sim N(0, \sigma^2), \mu_j \text{ 与 } \sigma^2 \text{ 未知, 且各 } \epsilon_{ij} \text{ 相互独立.} \end{cases} \quad (1)$$

上式称为单因素试验方差分析的数学模型.

若记 $n = \sum_{j=1}^r n_j$, $\mu = \frac{1}{n} \sum_{j=1}^r n_j \mu_j$ 称总均值, $\alpha_j = \mu_j - \mu \quad j=1, 2, \dots, r$. 则(1)式变形为

$$\begin{cases} X_{ij} = \mu + \alpha_j + \epsilon_{ij}, \\ \sum_{j=1}^r n_j \alpha_j = 0, \\ \epsilon_{ij} \sim N(0, \sigma^2), \text{ 各 } \epsilon_{ij} \text{ 相互独立.} \end{cases} \quad (2)$$

检验因素 A 的各水平的效应是否有显著差异,对(1)、(2)来说就是检验假设

$$H_0: \mu_1 = \mu_2 = \dots = \mu_r, H_1: \mu_1, \mu_2, \dots, \mu_r \text{ 不全相等},$$

或 $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0, H_1: \alpha_1, \alpha_2, \dots, \alpha_r \text{ 不全为零}.$

2. 统计分析

$$\text{记 } \bar{X} = \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^{n_j} X_{ij}, \bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij},$$

则有平方和分解公式 $S_T = S_A + S_E$.

$$\text{其中 } S_T = \sum_{j=1}^r \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 = \sum_{j=1}^r \sum_{i=1}^{n_j} X_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^r \sum_{i=1}^{n_j} X_{ij} \right)^2.$$

称为总离差平方和.

$$S_A = \sum_{j=1}^r n_j (\bar{X}_j - \bar{X})^2 = \sum_{j=1}^r \frac{1}{n_j} \left(\sum_{i=1}^{n_j} X_{ij} \right)^2 - \frac{1}{n} \left(\sum_{j=1}^r \sum_{i=1}^{n_j} X_{ij} \right)^2.$$

称为组间离差平方和,也称效应平方和.

$$S_E = \sum_{j=1}^r \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 = S_T - S_A.$$

称为组内离差平方和,也称误差平方和.

在 H_0 成立时,检验统计量

$$F = \frac{S_A / (r-1)}{S_E / (n-r)} = \frac{\bar{S}_A}{\bar{S}_E} \sim F(r-1, n-r).$$

对于显著性水平 α ,拒绝域为 $W = \{F > F_\alpha(r-1, n-r)\}$.

上述结果列成方差分析表如下:

方差来源	平方和	自由度	均方	F 值	F 的临界值
组间	S_A	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F = \frac{\bar{S}_A}{\bar{S}_E}$	$F_\alpha(r-1, n-4)$
组内	S_E	$n-r$	$\bar{S}_E = \frac{S_E}{n-r}$		
总和	S_T	$n-1$			

3. 具体计算过程

方差分析的基本步骤如下:

(1) 提出检验假设 $H_0: \mu_1 = \mu_2 = \dots = \mu_r, H_1: \mu_1, \mu_2, \dots, \mu_r \text{ 不全相等};$

(2) 计算 S_A, S_E 和 S_T , 并写出相应自由度, 进而计算 $\bar{S}_A, \bar{S}_E$;

(3) 计算 $F = \frac{\bar{S}_A}{\bar{S}_E}$, 并确定临界值 $F_\alpha(r-1, n-r)$;

(4) F 与临界值比较, 做出判断.

上述步骤有些计算常列表进行:

水平	因素 A 的水平						
观察值	A ₁	A ₂	...	A _j	...	A _r	
	x ₁₁	x ₂₁		x _{1j}		x _{1r}	
	x ₂₁	x ₂₂		x _{2j}		x _{2r}	
	⋮	⋮	...	⋮	...	⋮	
	x _{n₁1}	x _{n₂2}		x _{n_jj}		x _{n_rr}	
							∑
n _j	n ₁	n ₂	...	n _j	...	n _r	∑ _{j=1} ^r n _j
∑	∑ _{i=1} ^{n₁} x _{i1}	∑ _{i=1} ^{n₂} x _{i2}	...	∑ _{i=1} ^{n_j} x _{ij}	...	∑ _{i=1} ^{n_r} x _{ir}	∑ _{j=1} ^r ∑ _{i=1} ^{n_j} x _{ij}
(∑) ²	(∑ _{i=1} ^{n₁} x _{i1}) ²	(∑ _{i=1} ^{n₂} x _{i2}) ²	...	(∑ _{i=1} ^{n_j} x _{ij}) ²	...	(∑ _{i=1} ^{n_r} x _{ir}) ²	∑ _{j=1} ^r 1/n _j (∑ _{i=1} ^{n_j} x _{ij}) ²
∑ ²	∑ _{i=1} ^{n₁} x _{i1} ²	∑ _{i=1} ^{n₂} x _{i2} ²	...	∑ _{i=1} ^{n_j} x _{ij} ²	...	∑ _{i=1} ^{n_r} x _{ir} ²	∑ _{j=1} ^r ∑ _{i=1} ^{n_j} x _{ij} ²

$S_A = Q - P$ $S_E = R - Q$ $S_T = S_A + S_E$.

再由方差分析表做出判断.

8.1.2 典型题型解析

单因素方差分析解题过程如上所述.

例 8.1.1 某试验室对钢锭模进行选材试验时,将 4 种成分的生铁做成试样进行热疲劳测定.其方法是将试样加热到 700℃ 后投入 20℃ 的水中急冷,这样反复进行直至试样断裂,最后看试样经受的次数,显然经受的次数越多,质量就越好.

试验结果列于下面的表中,试检验 4 种生铁的试样抗疲劳性能是否有显著差异?

解 本例是单因素水平为 4 的不等重复试验, $n_1 = 7$ $n_2 = 5$ $n_3 = 8$ $n_4 = 6$ μ_1 μ_2 μ_3 μ_4 分别表示 4 种成分生铁的抗热疲劳性能(均值).我们要检验假设为

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $H_1: \mu_1, \mu_2, \mu_3, \mu_4$ 不全等.

列表计算如下:

	材料种类(因素水平)				
	1	2	3	4	
	160	158	146	151	
	161	164	155	152	
	165	164	160	153	
	168	170	162	157	
	170	175	164	160	
	172		166	168	
	180		174		
			182		∑
n _j	7	5	8	6	26
∑	1 176	831	1 309	941	4 257
(∑) ²	1 382 976	690 561	1 713 481	885 481	697 445.49
∑ ²	197 854	138 281	215 037	147 787	698 959

$P = 69\ 700\ 1.885$

$= Q$

$= R$

$$S_A = Q - P = 697\,445.49 - 697\,001.885 = 443.61,$$
$$S_E = R - Q = 698\,959 - 697\,445.49 = 1\,513.51,$$
$$S_T = R - P = 698\,959 - 697\,001.885 = 1\,957.115.$$

方差分析表为：

方差来源	平方和	自由度	均方	F 值	临界值	显著性
组间	443.6	3	147.87	2.15	$F_{0.05} = 3.65$	0.05
组内	1 513.51	22	68.78			
总和	1 957.115	25				

由于 $F = 2.15 < F_{0.05} = 3.65$,所以接受 H_0 ,认为四种生铁试样的热疲劳性能无显著差异.

例 8.1.2 一个年级有 3 个小班 ,他们进行了一次数学考试 ,现从各个班级随机地抽取了一些学生 ,成绩记录如下：

I	73	66	89	60	82	45	43	93	80	36	73	77			
II	88	77	78	31	48	78	91	62	51	76	85	96	74	80	56
III	68	41	79	59	56	68	91	53	71	79	71	15	87		

试在显著性水平 0.05 下检验各班级的平均分数有无显著差异. 设各个总体服从正态分布 ,且方差相等.

解 分别以 μ_1, μ_2, μ_3 表示 I、II、III 班的平均分数. 需检验假设

$$H_0 : \mu_1 = \mu_2 = \mu_3, H_1 : \mu_1, \mu_2, \mu_3 \text{ 不全相等}.$$

列表计算如下：

	不同班级(因素水平)				
	I	II	III		
	73 66 89 60 82 45 43 93 80 36 73 77	88 77 78 31 48 78 91 62 51 76 85 96 74 80 56	68 41 79 59 56 68 91 53 71 79 71 15 87	Σ	
n_j	12	15	13	40	
Σ	817	1 071	838	2 726	$P = 185\,776.9$
$(\Sigma)^2$	667 489	1 147 041	702 244	186 112.25	$= Q$
Σ^2	59 407	81 061	58 994	199 462	$= R$

$$S_A = Q - P = 186\,112.25 - 185\,776.9 = 335.35,$$
$$S_E = R - Q = 199\,462 - 186\,112.25 = 13\,349.75,$$
$$S_T = S_A + S_E = 13\,685.1.$$

方差分析表为：

方差来源	平方和	自由度	均方	F 值	临界值	显著性
组间	335.35	2	167.675	0.464 7	$F_{0.05} = 3.23$	0.05
组内	13 349.75	37	360.80			
总和	13 685.1	39				

因为 $F = 0.464 7 < F_{0.05} = 3.23$, 所以接受 H_0 .

例 8.1.3 在某单因素方差分析中 , 已知 $S_E = 204$, $S_T = 302.7$, $n - 1 = 25$, $\bar{S}_A = 24.675$, 求 S_A , S_A 自由度 , S_E 自由度及 F 值 .

解 由 $S_T = S_A + S_E$, 得 $S_A = S_T - S_E = 302.7 - 204 = 98.7$.

由 $\bar{S}_A = \frac{S_A}{r-1}$, 得 $r-1 = \frac{S_A}{\bar{S}_A} = \frac{98.7}{24.675} = 4$, 即 S_A 自由度为 4 .

此时 $r = 5$, 由 $n - 1 = 25$, 得 $n = 26$, $n - r = 26 - 5 = 21$, 即 S_E 自由度为 21 .

$$\bar{S}_E = \frac{S_E}{n-r} = \frac{204}{21} = 9.713 , F = \frac{\bar{S}_A}{\bar{S}_E} = \frac{24.675}{9.713} = 2.540 .$$

同步练习 8.1

1. 单因素方差分析中 , 检验统计量为 $F = \frac{\bar{S}_A}{\bar{S}_E}$, 则 F 服从分布().

A. $F(r, n)$ B. $F(r-1, n-r)$ C. $F(r-1, n-1)$ D. $F(n-r, n)$

2. 把一批同种纱线袜放在不同温度的水中洗涤 , 水温分为 6 个水平 , 每个水平下各洗 4 只袜子 , 袜子的收缩率以百分数记 , 其值如下表 . 试按显著水平为 0.05 和 0.01 判断不同洗涤水温对袜子的收缩率是否有显著影响 ?

水温	30℃	40℃	50℃	60℃	70℃	80℃
试号						
1	4.3	6.1	10	6.5	9.3	9.5
2	7.8	7.3	4.8	8.3	8.7	8.8
3	3.2	4.2	5.4	8.6	7.2	11.4
4	6.5	4.1	9.6	8.2	10.1	7.8

3. 用三种不同材料小球测定引力常数 , 实验结果如下表 , 试在 $\alpha = 0.01$ 下检验不同小球对引力常数的测定有无显著影响 ?

单位 : $10^{-11} \text{N} \cdot \text{m}^2 / \text{kg}^2$

铂	6.661	6.661	6.667	6.667	6.664
金	6.683	6.681	6.676	6.678	6.679
玻璃	6.678	6.671	6.675	6.672	6.674

同步练习 8.1 参考答案

1. B

2. 本题检验假设 :

$H_0: \mu_1 = \mu_2 = \dots = \mu_6$, $H_1: \mu_1, \mu_2, \dots, \mu_6$ 不全相等.

由题所给数据列表计算:

试号 \ 温度	因素水平								
	30℃	40℃	50℃	60℃	70℃	80℃			
1	4.3	6.1	10	6.5	9.3	9.5	Σ		
2	7.8	7.3	4.8	8.3	8.7	8.8			
3	3.2	4.2	5.4	8.6	7.2	11.4			
4	6.5	4.1	9.6	8.2	10.1	7.8			
n_j	4	4	4	4	4	4	24	$P = 1\ 315.72$	
Σ	21.8	21.7	29.8	31.6	35.3	37.5	177.7		
$(\sum)^2$	475.24	470.89	888.04	998.56	1 246.09	1 406.25	1 371.26		= Q
Σ ²	131.82	124.95	244.36	252.34	316.03	358.49	1 427.99		= R

$$S_A = Q - P = 1\ 371.26 - 1\ 315.72 = 55.54,$$

$$S_E = R - Q = 1\ 427.99 - 1\ 371.26 = 56.73,$$

$$S_T = S_A + S_E = 112.27.$$

S_T 的自由度为 $n - 1 = 24 - 1 = 23$,

S_A 的自由度为 $r - 1 = 6 - 1 = 5$,

S_E 的自由度为 $n - r = 24 - 6 = 18$.

$$\bar{S}_A = \frac{S_A}{r - 1} = 11.108, \bar{S}_E = \frac{S_E}{n - r} = 3.152.$$

方差分析表为:

方差来源	平方和	自由度	均方	F 值	临界值	显著性
组间	55.54	5	11.108	3.524	$F_{0.05} = 2.77$	0.05
组内	56.73	18	3.152		$F_{0.01} = 4.25$	0.01
总和	112.27	23				

由于 $2.77 < 3.524 < 4.25$, 所以拒绝 H_0 .

3. 把各小球测得的引力常数总体均值分别记为 μ_1, μ_2, μ_3 , 由题意需检验假设

$H_0: \mu_1 = \mu_2 = \mu_3$, $H_1: \mu_1, \mu_2, \mu_3$ 不全相等.

由题所给数据算得

$$S_T = \sum_{j=1}^r \sum_{i=1}^{n_j} X_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^r \sum_{i=1}^{n_j} X_{ij} \right)^2 = 705.94 \times 10^{-6},$$

$$S_A = \sum_{j=1}^r \frac{1}{n_j} \left(\sum_{i=1}^{n_j} X_{ij} \right)^2 - \frac{1}{n} \left(\sum_{j=1}^r \sum_{i=1}^{n_j} X_{ij} \right)^2 = 565.11 \times 10^{-6},$$

$$S_E = S_T - S_A = 140.83 \times 10^{-6}.$$

S_T 的自由度为 $n - 1 = 16 - 1 = 15$,

S_A 的自由度为 $r - 1 = 3 - 1 = 2$,

S_E 的自由度为 $n - r = 16 - 3 = 13$.

$$F = \frac{\frac{\bar{S}_A}{\bar{S}_E} = \frac{\frac{S_A}{n-1}}{\frac{S_E}{n-r}} = \frac{282.555 \times 10^{-6}}{10.833 \times 10^{-6}} = 26.08.$$

由 $\alpha = 0.01$ 查表得 $F_{\alpha}(r-1, m-r) = F_{0.01}(2, 13) = 6.70$.

方差分析表为

方差来源	平方和	自由度	均方	F 值	临界值	显著性
组间	565.11×10^{-6}	2	282.555×10^{-6}	26.08	$F_{0.01} = 6.70$	0.01
组内	140.83×10^{-6}	13	10.833×10^{-6}			
总和	705.94×10^{-6}	15				

由于 $F = 26.08 > 6.70 = F_{0.01}(2, 13)$, 所以拒绝 H_0 , 认为不同小球对引力常数测定有高度显著的影响.

注 本题解法中直接利用 S_A 、 S_T 定义来计算的, 没有采用列表计算的方法, 当然也可采用列表计算的方法(如例 8.1.1).

8.2 双因素方差分析

8.2.1 知识要点

1. 数学模型

设因素 A 有 r 个水平 $A_1, A_2, \dots, A_r$, 因素 B 有 s 个水平 $B_1, B_2, \dots, B_s$. 对因素 A, B 的各种水平的每一对组合 $(A_i, B_j)(i = 1, 2, \dots, r; j = 1, 2, \dots, s)$ 进行一次试验, 并假定试验是相互独立的. 试验结果如下表:

样本 因素 A \ 因素 B						
	B_1	B_2	...	B_j	...	B_s
A_1	X_{11}	X_{12}	...	X_{1j}	...	X_{1s}
A_2	X_{21}	X_{22}	...	X_{2j}	...	X_{2s}
A_i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{is}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_r	X_{r1}	X_{r2}	...	X_{rj}	...	X_{rs}

设 $X_{ij} \sim N(\mu_{ij}, \sigma^2)$, μ_{ij}, σ^2 未知, 各 X_{ij} 相互独立. 有

$$\begin{cases} X_{ij} = \mu_{ij} + \epsilon_{ij}, & i = 1, 2, \dots, r; j = 1, 2, \dots, s, \\ \epsilon_{ij} \sim N(0, \sigma^2), & \text{且各 } \epsilon_{ij} \text{ 相互独立.} \end{cases} \tag{3}$$

引入记号

$$\begin{cases} \mu = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \mu_{ij}, \\ \mu_i = \frac{1}{s} \sum_{j=1}^s \mu_{ij}, \\ \mu_j = \frac{1}{r} \sum_{i=1}^r \mu_{ij}, \\ \alpha_i = \mu_i - \mu, \quad i=1, 2, \dots, r, \\ \beta_j = \mu_j - \mu, \quad j=1, 2, \dots, s. \end{cases}$$

代入(3)从而有

$$\begin{cases} X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i=1, 2, \dots, r; j=1, 2, \dots, s, \\ \varepsilon_{ij} \sim N(0, \sigma^2), \text{ 且各 } \varepsilon_{ij} \text{ 相互独立.} \end{cases}$$

要判断 A、B 的影响是否显著, 需检验假设

$$\begin{cases} H_{0A}: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0; H_{1A}: \alpha_1, \alpha_2, \dots, \alpha_r \text{ 不全为零;} \\ H_{0B}: \beta_1 = \beta_2 = \dots = \beta_s = 0; H_{1B}: \beta_1, \beta_2, \dots, \beta_s \text{ 不全为零.} \end{cases}$$

2. 统计分析

记

$$\begin{cases} \bar{X} = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s X_{ij}, \\ \bar{X}_{i\cdot} = \frac{1}{s} \sum_{j=1}^s X_{ij}, \quad i=1, 2, \dots, r, \\ \bar{X}_{\cdot j} = \frac{1}{r} \sum_{i=1}^r X_{ij}, \quad j=1, 2, \dots, s. \end{cases}$$

又记

$$\begin{cases} T_{i\cdot} = \sum_{j=1}^s X_{ij}, \\ T_{\cdot j} = \sum_{i=1}^r X_{ij}, \\ T = \sum_{i=1}^r \sum_{j=1}^s X_{ij}, \end{cases}$$

则有平方和分解公式

$$S_T = S_E + S_A + S_B,$$

其中

$$S_T = \sum_{i=1}^r \sum_{j=1}^s (X_{ij} - \bar{X})^2 = \sum_{i=1}^r \sum_{j=1}^s X_{ij}^2 - \frac{T^2}{rs},$$

$$S_A = s \sum_{i=1}^r (\bar{X}_{i\cdot} - \bar{X})^2 = \frac{1}{s} \sum_{i=1}^r T_{i\cdot}^2 - \frac{T^2}{rs},$$

$$S_B = r \sum_{j=1}^s (\bar{X}_{\cdot j} - \bar{X})^2 = \frac{1}{r} \sum_{j=1}^s T_{\cdot j}^2 - \frac{T^2}{rs},$$

$$S_E = S_T - S_A - S_B.$$

可以证明：

- (1) $S_E/\sigma^2 \sim \chi^2((r-1)(s-1))$;
(2) 当 H_{0A} 成立时 $S_A/\sigma^2 \sim \chi^2(r-1)$, 且 S_A 与 S_E 独立 ,

$$F_A = \frac{S_A/(r-1)}{S_E/((r-1)(s-1))} \sim F(r-1, (r-1)(s-1)) ;$$

- (3) 当 H_{0B} 成立时 $S_B/\sigma^2 \sim \chi^2(s-1)$, 且 S_B 与 S_E 独立 ,

$$F_B = \frac{S_B/(s-1)}{S_E/((r-1)(s-1))} \sim F(s-1, (r-1)(s-1)) .$$

F_A, F_B 为检验 H_{0A}, H_{0B} 的统计量 , 对给定显著性水平 α , 当 $F_A \geq F_{\alpha}(r-1, (r-1)(s-1))$ 时 , 拒绝 H_{0A} ; 当 $F_B \geq F_{\alpha}(s-1, (r-1)(s-1))$ 时 , 拒绝 H_{0B} .

将上述分析结果列成方差分析表为：

方差来源	平方和	自由度	均方	F 值	临界值	显著性
因素 A	S_A	$r-1$	$\bar{S}_A = S_A/(r-1)$	$F_A = \frac{\bar{S}_A}{S_E}$	$F_{\alpha}(r-1, (r-1)(s-1))$	α
因素 B	S_B	$s-1$	$\bar{S}_B = S_B/(s-1)$	$F_B = \frac{\bar{S}_B}{S_E}$	$F_{\alpha}(s-1, (r-1)(s-1))$	
误差	S_E	$(r-1)(s-1)$	$\bar{S}_E = S_E/((r-1)(s-1))$			
总和	S_T	$rs-1$				

8.2.2 典型题型解析

双因素方差分析解题要点为：根据题意写出待检假设，再计算相关统计量的值，在查出临界值的基础上做出判断。解此类题常列表进行。

例 8.2.1 为研究各种不同的土质对两种钢管的腐蚀，在土中埋了八年后测得被腐蚀的质量列于下表。试在显著水平 $\alpha = 0.05$ 下检验不同土质对钢管腐蚀的差异性，以及不同钢管腐蚀情况的差异性。

单位：kg

A		涂沿钢管	裸露钢管
B			
	细砂土	0.18	1.70
	砾砂土	0.08	0.21
	淤泥	0.61	1.21
	黏土	0.44	0.89
	沼泽地	0.77	0.86
	碱土	1.27	2.64

解 本题是双因素无交互作用试验的方差分析问题。以 $\alpha_1, \alpha_2, \dots, \alpha_6$ 表示土质因素 A 的各水平的效应， β_1, β_2 表示涂层因素 B 各水平的效应。

检测假设 $H_{0A} : \alpha_1 = \alpha_2 = \dots = \alpha_6 = 0, H_{1A} : \alpha_1, \alpha_2, \dots, \alpha_6$ 不全为零；

$H_{0B} : \beta_1 = \beta_2 = 0, H_{1B} : \beta_1, \beta_2$ 不全为零。

由题设 $r = 6, s = 2$ ，列表计算如下：

续表

因素 A \ 因素 B	涂铅钢管 1	裸露钢管 2	$T_{i\cdot}$	$T_{i\cdot}^2$
细砂土 1	0.18	1.70	1.88	3.534 4
砾砂土 2	0.08	0.21	0.29	0.084 1
淤泥 3	0.61	1.21	1.82	3.312 4
黏土 4	0.44	0.89	1.33	1.768 9
沼泽地 5	0.77	0.86	1.63	2.656 9
碱土 6	1.27	2.64	3.91	15.288 1
$T_{\cdot j}$	3.35	7.51	$T = \sum_{j=1}^s T_{\cdot j} = 10.86, \sum_{j=1}^s T_{\cdot j}^2 = 67.622 6$	
$T_{\cdot j}^2$	11.222 5	56.400 1	$\sum_{i=1}^r T_{i\cdot}^2 = 26.644 8, \sum_{j=1}^s \sum_{i=1}^r Z_{ij}^2 = 15.709 8$	

由此算得

$$S_T = \sum_{i=1}^r \sum_{j=1}^s Z_{ij}^2 - \frac{T^2}{rs} = 15.709 8 - \frac{1}{12} \times 10.86^2 = 5.881 5,$$

$$S_A = \frac{1}{s} \sum_{i=1}^r T_{i\cdot}^2 - \frac{T^2}{rs} = \frac{1}{2} \times 26.644 8 - \frac{1}{12} \times 10.86^2 = 3.494 1,$$

$$S_B = \frac{1}{r} \sum_{j=1}^s T_{\cdot j}^2 - \frac{T^2}{rs} = \frac{1}{6} \times 67.622 6 - \frac{1}{12} \times 10.86^2 = 1.442 1,$$

$$S_E = S_T - S_A - S_B = 5.881 5 - 3.494 1 - 1.442 1 = 0.945 3.$$

 S_A 的自由度 $r - 1 = 5$, S_B 的自由度 $s - 1 = 1$, S_E 的自由度 $(r - 1)(s - 1) = 5$.

于是得到

$$F_A = \frac{S_A / 5}{S_E / 5} = \frac{0.698 82}{0.189 06} = 3.696,$$

$$F_B = \frac{S_B / 1}{S_E / 5} = \frac{1.442 1}{0.189 06} = 0.762 8.$$

方差分析表为：

方差来源	平方和	自由度	均方	F 值	临界值	显著性
A	3.494 1	5	0.698 82	$F_A = 3.696$	$F_{0.95}(5, 5) = 5.05$	0.05
B	1.442 1	1	1.442 1	$F_B = 7.628$	$F_{0.95}(1, 5) = 6.61$	0.05
误差	0.945 3	5	0.189 06			
总和	$S_T = 5.8815$	11				

由于 $F_A = 3.696 < 5.05 = F_{0.95}(5, 5)$, 所以因素 A (即土质) 对钢管的腐蚀性影响是不显著的. 又 $F_B = 7.628 > 6.61 = F_{0.95}(1, 5)$, 所以因素 B (即钢管涂层) 对其腐蚀性的影响是显著的.

例 8.2.2 设 3 种配料方案分别与 5 种不同的硫化时间 (单位: s) 搭配起来生产橡胶, 测得橡胶的抗断强度 (单位: kg/cm^2) 如下表:

配料A \ 硫化时间 B					
	B ₁	B ₂	B ₃	B ₄	B ₅
A ₁	152	158	149	143	126
A ₂	146	163	150	151	137
A ₃	125	144	138	130	121

经计算得 $S_A = 878.8$, $S_B = 1\,140$, $S_E = 163.6$, $S_T = 218.4$, 在显著水平 $\alpha = 0.01$ 下分别检验各种配料方案与各种硫化时间对橡胶抗断强度是否有显著影响.

解 计算得 S_A 自由度 $r - 1 = 3 - 1 = 2$,
 S_B 自由度 $s - 1 = 5 - 1 = 4$,
 S_E 自由度 $(r - 1)(s - 1) = 8$.

均方 : $\bar{S}_A = \frac{S_A}{r - 1} = \frac{878.8}{2} = 439.4$,
 $\bar{S}_B = \frac{S_B}{s - 1} = \frac{1\,140}{4} = 285$,
 $\bar{S}_E = \frac{S_E}{(r - 1)(s - 1)} = \frac{163.6}{8} = 20.45$.

F 值 : $F_A = \frac{\bar{S}_A}{\bar{S}_E} = \frac{439.4}{20.45} = 21.49$,
 $F_B = \frac{\bar{S}_B}{\bar{S}_E} = \frac{285}{20.45} = 13.94$.

列出方差分析表为 :

方差来源	平方和	自由度	均方	F 值	临界值	显著性
A	878.8	2	439.4	$F_A = 21.49$	$F_{0.01}(2, 8) = 8.65$	0.01
B	1 140	4	285	$F_B = 13.94$	$F_{0.01}(4, 8) = 7.01$	0.01
误差	163.6	8	20.45			
总和	2 182.4	14				

由于 $F_A = 21.49 > 8.65 = F_{0.01}(2, 8)$, 所以 , 可以认为各种配料方案下生产橡胶的抗断强度有显著差异 ; $F_B = 13.94 > 7.01 = F_{0.01}(4, 8)$, 所以可认为各种硫化时间下生产橡胶的抗断强度有显著差异.

同步练习 8.2

1. 试验某种钢不同的含钢量在各种温度下的冲击值 (kg · m / cm²). 其实测数据如下表 , 试检验差异性是否显著 ?

A 试验温度 \ B 铜含量 %	0.2%	0.4%	0.8%	$T_{i \cdot}$	$T_{i \cdot}^2$
20℃	10.6	11.6	14.5	36.7	1 346.9
0℃	7.0	11.1	13.3	31.4	985.9
- 20℃	4.2	6.8	11.5	22.5	506.3
- 40℃	4.2	6.3	8.7	19.2	368.6
$T_{\cdot j}$	26	35.8	48	$T = 109.8$	
$T_{\cdot j}^2$	676	1 281.6	2 304		

2. 为了研究蒸馏水的 pH 值和硫酸铜溶液浓度对化验血清中的白蛋白与球蛋白的影响 ,

对蒸馏水的 pH 值(A)取了四个不同水平,对硫酸的浓度(B)取了三个不同水平,在不同水平组合(A_i, B_j)下,各测一次白蛋白与球蛋白之比,将其结果列于下表.试在 $\alpha = 0.05$ 下检验两个因素对化验结果有无显著差异.

$A \backslash B$	B_1	B_2	B_3	$T_{i\cdot}$
A_1	3.5	2.3	2.0	7.8
A_2	2.6	2.0	1.9	6.5
A_3	2.0	1.5	1.2	4.7
A_4	1.4	0.8	0.3	2.5
$T_{\cdot j}$	9.5	6.6	5.4	

同步练习 8.2 参考答案

1. 本题中 $r = 4, s = 3$.

$$S_T = \sum_{i=1}^r \sum_{j=1}^s X_{ij}^2 - \frac{T^2}{rs} (10.6^2 + \dots + 8.7^2) - \frac{109.8^2}{12} = 130.75,$$

$$S_A = \frac{1}{s} \sum_{i=1}^r T_{i\cdot}^2 - \frac{T^2}{rs} = \frac{1}{3} (1\ 346.9 + \dots + 368.6) - \frac{109.8^2}{12} = 64.56,$$

$$S_B = \frac{1}{r} \sum_{j=1}^s T_{\cdot j}^2 - \frac{T^2}{rs} = \frac{1}{4} (676 + \dots + 2\ 304) - \frac{109.8^2}{12} = 60.73,$$

$$S_E = S_T - S_A - S_B = 130.75 - 64.56 - 60.73 = 5.46.$$

S_A 自由度 $r - 1 = 3$,

S_B 自由度 $s - 1 = 2$,

S_E 自由度 $(r - 1)(s - 1) = 6$.

于是

$$F_A = \frac{S_{A/2}}{S_{E/6}} = \frac{21.52}{0.91} = 23.65,$$

$$F_B = \frac{S_{B/2}}{S_{E/6}} = \frac{30.37}{0.91} = 33.37.$$

误差来源	平方和	自由度	均方	F 值	临界值	显著性
A 温度作用	$S_A = 64.56$	3	21.52	$F_A = 23.65$	$F_{0.01}(3, 6) = 9.78$	0.01
B 铜含量作用	$S_B = 60.73$	2	30.37	$F_B = 33.37$	$F_{0.01}(2, 6) = 10.92$	0.01
误差	$S_E = 5.46$	6	0.91			
总和	$S_T = 130.75$	11				

由于 $F_A = 23.65 > 9.78 = F_{0.01}(3, 6)$, 所以钢随着试验温度的升高, 冲击值也越来越大;
 $F_B = 33.37 > 10.92 = F_{0.01}(2, 6)$, 所以钢随着含铜量的增加冲击值提高也是显著的.

2. 设 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 表示蒸馏水 pH 值四个水平效应; $\beta_1, \beta_2, \beta_3$ 表示硫酸浓度三个水平的效应. 待检假设

$H_{0A}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4, H_{1A}: \alpha_1, \alpha_2, \alpha_3, \alpha_4$ 不全相等

$H_{0B}: \beta_1 = \beta_2 = \beta_3, H_{1B}: \beta_1, \beta_2, \beta_3$ 不全相等

通过计算得方差分析表

方差来源	平方和	自由度	均方	F 值	临界值	显著性
A	$S_A = 5.29$	3	1.76	$F_A = 40.9$	$F_{0.05}(3, 6) = 4.8$	0.05
B	$S_B = 2.22$	2	1.11	$F_B = 25.8$	$F_{0.05}(2, 6) = 5.1$	0.05
误差	$S_E = 0.26$	6	0.043			
总和	S_T	11				

由于 $F_A = 40.9 > F_{0.05}(3, 6)$, 拒绝 H_{0A} , 接受 H_{1A} , 即因素 A 对化验结果有显著影响; 由 $F_B = 25.8 > F_{0.05}(2, 6)$, 拒绝 H_{0B} , 即因素 B 不同对化验结果有显著影响。

疑难解析

1. 方差分析的理论依据是什么?

答 方差分析有单因素方差分析、双因素方差分析及多因素方差分析等种类, 它们都基于以下的理论假定。

(1) 正态性假定. 在正态性假定情形下, 数据 x_{ij} 认为取自 $N(\mu, \sigma^2)$, 由此求得的各种离差平方和(如比值 S_T / σ^2)服从 χ^2 分布, 于是定义 F 分布函数. 若没有正态假定, 则没有 χ^2 分布, 从而也就没有 F 分布与统计推断。

(2) 方差齐性假定. 假定数据 x_{ij} 来自方差的 σ^2 的正态总体, 只有这样才能在相同的条件 (σ^2 相等) 下来分析问题. 考察指标的变化, 才可建立统计假设 H_0 , 才有方差分析检验。

(3) 线性假定. 线性假定指数据 x_{ij} 的取得仅通过线性运算, 因而才可以把数据 x_{ij} 当作线性模型处理, 也才可以施行方差分析。

大数定律和中心极限定理确立了正态性假定的合理性、数据的线性假设也符合实际易于成立. 方差齐性假设不易确立, 对于不同组数据, 很难保持方差齐性, 所以常用数据的变换来实现. 有些实际问题的方差分析前一般需进行方差齐性检验. 在方差分析中, 三个假定缺一不可, 否则方差分析就失去了依据。

2. 在实际问题中, 怎样确定方差分析中所考虑因素的个数?

答 在实际问题中, 影响指标的因素往往是较多的, 这就要求根据问题的性质, 对问题的了解和研究的规模确定因素, 即确定因素是什么, 并将其分成几个水平, 区分水平的依据. 在实际研究中, 为方便和突出结论, 一般只让一个或两个因素变化, 而将其他因素固定下来. 这样观察和分析指数的变化和因素的影响, 即为单因素或双因素方差分析. 若这样的因素比较多时, 这样的分析称为多因素(元)方差分析, 其数学处理比较繁琐。

3. 方差分析与 t 检验、 F 检验有什么关系?

答 方差分析是检验几个总体的平均数来自同一正态总体的可信程度。

方差分析是由 t 检验发展来的. 它们都是检验几个总体平均数是否来自同一正态总体的可信程序. 当因素的水平为 2 时, 方差分析与 t 检验是一致的. 因素水平(单因素)为 2 时, F 统计量的第一自由度是 1, F 分布表中的 F 值与 t 分布表中的 t 值存在平方根关系, 即 $t_{\frac{\alpha}{2}}(1) = \sqrt{F_{\alpha}(1, 1)}$. 说明 t 函数的平方是一个 F 函数. 但 t 检验只适合水平为 2 的情形, 而方差分析

适合水平大于等于 2 的情形.

F 检验法是检验两个总体的方差来自同一正态总体的可信程度,两者的出发点是不同的.方差分析中,规定 $S_A/(r-1)$ 是第一样本作为分子, $S_E/(n-r)$ 是第二样本作为分母,而 F 检验没有规定哪个样本为分子或分母,通常是算出数值后,以数值大的作为第一样本放在统计量的分子上,这是与方差分析的又一重要差别.在方差分析中,当 H_0 成立时, $[S_A/(r-1)]/[S_E/(n-r)]$ 服从 F 分布,当 H_0 不成立时,其分布图像较之偏右移,因而在方差分析时,通常采用单侧检验.在 F 检验中, H_0 成立时,两样本的方差比服从 F 分布; H_0 不成立时,分布图形可能偏右也可能偏左.

本章常考题型归纳

考点	常见题型	例题
方差分析基本概念	方差分析数学模型认知	
单因素方差分析	单因素方差分析的统计分析(S_A 、 S_E 和 S_T)	例 8.1.3
	单因素方差分析	例 8.1.1,例 8.1.2
双因素方差分析	双因素方差分析数学模型认知	
	双因素方差分析的统计分析	
	双因素方差分析	例 8.2.1,例 8.2.2

本章知识结构网络



检测题(一)

1. 对单因素方差分析模型

$$\begin{cases} X_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i=1, 2, \dots, m; \quad j=1, 2, \dots, r \\ \sum_{j=1}^r n_j \alpha_j = 0, \\ \epsilon_{ij} \sim N(0, \sigma^2), \text{各 } \epsilon_{ij} \text{ 相互独立.} \end{cases}$$

求 $E(S_A)$

2. 将抗生素注入人体会产生抗生素与血浆蛋白质结合的现象,以致减少了药效,下表列出 5 种常用的抗生素注入到牛的体内时,抗生素与血浆蛋白质结合的百分比,试在水平 $\alpha = 0.05$ 下检验这些百分比的均值有无显著的差异. 设备总体服从正态分布,且方差相同.

青霉素	四环素	链霉素	红霉素	氯霉素
29.6	27.3	5.8	21.6	29.2
24.3	32.6	6.2	17.4	32.8
28.5	30.8	11.0	18.3	25.0
32.0	34.8	8.3	19.0	24.2

3. 为了研究金属管的防腐蚀的功能,考虑了 4 种不同涂料涂层,将金属管埋设在 3 种不同性质的土壤中,经历了一定时间,测得金属管腐蚀的最大深度如下表所示(单位 mm):

涂层 A	土壤 B		
	1	2	3
1	1.63	1.35	1.27
2	1.34	1.30	1.22
3	1.19	1.14	1.27
4	1.30	1.09	1.32

设两因素间没有交互作用效应. 试取水平 $\alpha = 0.05$ 检验在不同涂层下腐蚀的最大深度的平均值有无显著差异,在不同土壤下腐蚀的最大深度的平均值有无显著差异.

检测题(一)参考答案

1. 略

2. 分别以 $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ 表示青霉素,四环素,链霉素,红霉素,氯霉素的百分比的均值. 需检验假设

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5, H_1: \mu_1, \mu_2, \mu_3, \mu_4, \mu_5 \text{ 不全相等.}$$

方差分析表如下:

方差来源	平方和	自由度	均方	F 值	临界值	显著性
组间	1 480.823	4	370.21	40.91	$F_{0.05}(4, 15) = 3.06$	0.05
组内	1 35.823	15	9.05			
总和	1 616.646	19				

由于 $F = 40.91 > F_{0.05}(4, 15) = 3.06$, 所以拒绝 H_0 , 即差异显著.

3. 分别以 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 表示 4 种不同的涂料涂层下腐蚀的最大深度的均值, 以 $\beta_1, \beta_2, \beta_3$ 表示三种不同土壤下腐蚀的最大深度的均值, 检验假设

$$H_{0A}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4, H_{1A}: \alpha_1, \alpha_2, \alpha_3, \alpha_4 \text{ 不全相等,}$$

$$H_{0B}: \beta_1 = \beta_2 = \beta_3, H_{1B}: \beta_1, \beta_2, \beta_3 \text{ 不全相等.}$$

方差分析表为

方差来源	平方和	自由度	均方	F 值	临界值	显著性
A	0.080 7	3	0.026 9	$F_A = \frac{\bar{S}_A}{\bar{S}_E}$	$F_{0.05}(3, 6) = 4.76$	0.05
B	0.043 4	2	0.021 7	$= 2.107 0$	$F_{0.05}(2, 6) = 5.14$	
误差	0.076 6	6	0.012 77	$F_B = \frac{\bar{S}_B}{\bar{S}_E}$		
总和	0.200 7	11		$= 1.699 7$		

因为 $F_A = 2.107 0 < 4.76 = F_{0.05}(3, 6)$, 所以接受 H_{0A} , 即涂层因素影响不显著; 因为 $F_B = 1.699 7 < 5.14 = F_{0.05}(2, 6)$, 所以接受 H_{0B} , 即土壤因素影响不显著.

检测题(二)

1. 设有三台同样规格的机器, 用来生产厚度为 $\frac{1}{4}$ cm 的铝板, 各台机器生产的产品平均厚度是否相同, 取样测得结果如表. 试在显著水平 $\alpha = 0.05$ 下检验差异显著性.

机号		1	2	3
试号				
1		0.236	0.257	0.258
2		0.238	0.253	0.264
3		0.248	0.255	0.259
4		0.245	0.254	0.267
5		0.243	0.261	0.262

2. 今有某种型号的电池三批, 它们分别是 A 、 B 、 C 三个工厂生产的, 为详比其质量, 各随机抽取 5 只电池为样品, 经试验得其寿命(小时)如下:

T		A	B	C
试号				
1		40	26	39
2		48	34	40
3		38	30	43
4		42	28	50
5		45	32	50

试在显著性水平 $\alpha = 0.05$ 下, 检验电池的平均寿命有无显著的差异. 若差异是显著的, 试求均值差 $\mu_A - \mu_B$, $\mu_A - \mu_C$ 及 $\mu_B - \mu_C$ 的置信度为 95% 的置信区间. 设各工厂生产的电池寿命服从同方差的正态分布.

3. 为了研究三种化学肥料和四种土壤对种子发芽率的影响, 经过一段实验后测得发芽率如下表所示. 假设肥料因素、土壤因素没有交互作用. 试取水平 $\alpha = 0.05$ 检验不同土壤下发芽率平均值有无显著差异, 在不同化肥下发芽率平均值有无显著差异.

化肥		A	B	C
土壤				
1		1.63	1.35	1.27
2		1.34	1.30	1.22
3		1.19	1.14	1.27
4		1.30	1.09	1.32

检测题(二)参考答案

1. 本题是单因素方差分析问题. 设 μ_1, μ_2, μ_3 表示各台机器生产的铝板平均厚度, 问题是在显著性水平 $\alpha = 0.05$ 下检验假设

$H_0: \mu_1 = \mu_2 = \mu_3, H_1: \mu_1, \mu_2, \mu_3$ 不全相等.

方差分析表为:

方差来源	平方和	自由度	均方	F 值	临界值	显著性
机器	1 053.33	2	526.665	32.92	$F_{0.05} = 3.89$	0.05
误差	192	12	15			
总和	1 245.33	14				

由于 $3.89 < 32.92$, 故拒绝 H_0 ,

2. $H_0: \mu_A = \mu_B = \mu_C, H_1: \mu_A, \mu_B, \mu_C$ 不全相等.

由数据, 同减 40, 然后计算,

方差分析表

方差来源	平方和	自由度	均方	F 值	临界值	显著性
因素	615.6	2	307.8	17.068 4	3.89	0.05
误差	216.4	12	18.033 3			
总和	832	14				

即认为电池的平均寿命有显著差异($F = 17.068 > 3.89$).

置信度为 0.95 的置信区间如下:

$\mu_A - \mu_B$ 的置信度为 0.95 的置信区间为 $(6.75, 18.45)$, $\mu_A - \mu_C$ 的置信区间为 $(-7.652, 4.052)$, $\mu_B - \mu_C$ 的置信区间为 $(-20.252, -8.548)$.

3. 分别以 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 表示 4 种土壤发芽率平均值, 以 b_1, b_2, b_3 表示三种化肥发芽率平均值, 检验假设为

$$\begin{cases} H_{01}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4, \\ H_{11}: \alpha_1, \alpha_2, \alpha_3, \alpha_4 \text{ 不全相等.} \end{cases}$$
$$\begin{cases} H_{02}: b_1 = b_2 = b_3, \\ H_{12}: b_1, b_2, b_3 \text{ 不全相等.} \end{cases}$$

方差分析表:

方差来源	平方和	自由度	均方	$F_{\text{比}}$
因素 A	0.080 7	3	0.026 9	$F_A = \frac{\bar{S}_A}{\bar{S}_E} = 2.107 0$
因素 B	0.043 4	2	0.021 7	
误差	0.076 6	6	0.012 77	$F_B = \frac{\bar{S}_B}{\bar{S}_E} = 1.699 7$
总和	6.207 7	11		

$F_{\alpha}(r-1, (r-1)(s-1)) = F_{0.05}(3, 6) = 4.76 > 2.107 0,$

$F_{\alpha}(s-1, (r-1)(s-1)) = F_{0.05}(2, 6) = 5.14 > 1.699 7.$

故在 $\alpha = 0.05$ 下, 接受 H_{01}, H_{02} , 即认为两种因素影响均不显著.

第 9 章 回归分析

9.1 一元线性回归

9.1.1 知识要点

1. 一元线性回归模型

设随机变量 Y (因变量) 与普通变量 x (自变量) 存在着相关关系. 对于 x 的每一确定值, Y 有它的分布. 若 Y 的数学期望存在, 则期望值随 x 的取值而定, 即 Y 的数学期望是 x 的函数, 记为 $\mu(x)$. 这里 $\mu(x)$ 称为 Y 关于 x 的回归函数.

一元线性回归是研究 $\mu(x)$ 为 x 的线性函数即 $\mu(x) = a + bx$ 的情况. 一元线性回归模型为

$$\left. \begin{aligned} Y &= a + bx + \varepsilon, \\ \varepsilon &\sim N(0, \sigma^2), \end{aligned} \right\} \quad (1)$$

其中 a, b 及 σ 都不依赖于 x , 且 a, b, σ^2 均未知.

2. 回归方程

若由样本得到(1)中参数 a, b 的估计 $\hat{a}, \hat{b}$, 对于给定的 x , 定义 $\hat{y} = \hat{a} + \hat{b}x$ 作为 $\mu(x) = a + bx$ 的估计, 并且称方程

$$\hat{y} = \hat{a} + \hat{b}x \quad (2)$$

为 y 关于 x 的线性回归方程或回归方程, 其图像称为回归直线, $\hat{a}, \hat{b}$ 称为回归系数.

3. 回归系数的估计

取 x 的 n 个不全相同的值 $x_1, x_2, \dots, x_n$ 做独立试验, 得到样本值 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, a, b 的极大似然估计值可按下面方法计算. 引入记号

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2,$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2,$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right),$$

则 a, b 的极大似然估计值为

$$\hat{b} = \frac{S_{xy}}{S_{xx}}, \quad \hat{a} = \frac{1}{n} \sum_{i=1}^n y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \hat{b} = \bar{y} - \bar{x} \hat{b}.$$

4. 误差 ε 的方差 $D(\varepsilon) = \sigma^2$ 的无偏估计

$$\text{记 } Q_E = \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i)^2 = S_{yy} - \hat{b}S_{xy}.$$

称 Q_E 为残差平方和, σ^2 的无偏估计为

$$\hat{\sigma}^2 = \frac{Q_E}{n-2}.$$

5. 回归系数 b 的置信区间

b 的一个置信水平为 $1-\alpha$ 的置信区间为

$$\left(\hat{b} - t_{\frac{\alpha}{2}}(n-1) \frac{\hat{\sigma}}{\sqrt{S_{xx}}}, \hat{b} + t_{\frac{\alpha}{2}}(n-1) \frac{\hat{\sigma}}{\sqrt{S_{xx}}} \right),$$

其中 $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$.

6. 回归函数值 $\mu(x_0)$ 的置信区间

$\mu(x)$ 在点 x_0 处的函数值 $\mu(x_0)$ 的置信水平 $1-\alpha$ 的置信区间为

$$\left[\hat{a} + \hat{b}x_0 - t_{\frac{\alpha}{2}}(n-2) \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{a} + \hat{b}x_0 + t_{\frac{\alpha}{2}}(n-2) \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right].$$

7. 回归方程的显著性检验

线性假设 $H_0: b=0$, $H_1: b \neq 0$ 的显著性检验, 若经检验拒绝 H_0 认为回归效果是显著的; 否则, 则认为回归效果不显著, 此时不宜用线性回归模型, 需另行研究.

主要有三种检验方法:

① t 检验法. 检验统计量为

$$t = \frac{\hat{b} \sqrt{n[x^2 - (\bar{x})^2]}}{\sqrt{S_E} \sqrt{n-2}},$$

对显著性水平 α , H_0 的拒绝域为

$$|t| \geq t_{\frac{\alpha}{2}}(n-2).$$

② F 检验法. 检验统计量为

$$F = \frac{S_R}{S_E(n-2)}, \text{ 其中 } S_R = \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{n(\overline{xy} - \bar{x} \cdot \bar{y})^2}{x^2 - (\bar{x})^2},$$

对显著性水平 α , H_0 的拒绝域为

$$F \geq F_{\alpha}(1, n-2).$$

③ 相关系数检验法. 检验统计量

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{x^2 - (\bar{x})^2} \sqrt{y^2 - (\bar{y})^2}},$$

对显著性水平 α , 可查表得临界值 C , 当 $|r| \geq C$ 则拒绝 H_0 .

8. Y 在 x_0 处的观察值 Y_0 的预测值以及预测区间

以 x_0 处的回归值 $\hat{y}_0 = \hat{a} + \hat{b}x_0$ 作为 Y 在 x_0 处的观察值 $Y_0 = a + bx_0 + \varepsilon_0$ 的预测值.

Y_0 的置信水平为 $1-\alpha$ 的预测区间为

$$\left[\hat{a} + \hat{b}x_0 - t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \quad \hat{a} + \hat{b}x_0 + t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right].$$

9. 可化为线性回归的例子

$$\textcircled{1} y = a + b \sin x + \varepsilon, \varepsilon \sim N(0, \sigma^2),$$

若令 $u = \sin x$, 可化为 $y = a + bu + \varepsilon, \varepsilon \sim N(0, \sigma^2)$.

$$\textcircled{2} y = a + \frac{b}{x} + \varepsilon, \varepsilon \sim N(0, \sigma^2),$$

若令 $u = \frac{1}{x}$, 可化为 $y = a + bu + \varepsilon, \varepsilon \sim N(0, \sigma^2)$.

$$\textcircled{3} \varphi(y) = a + bx + \varepsilon, \varepsilon \sim N(0, \sigma^2),$$

若令 $z = \varphi(y)$, 得 $z = a + bx + \varepsilon, \varepsilon \sim N(0, \sigma^2)$.

$$\textcircled{4} y = ax^b \varepsilon_0, x > 0, \ln \varepsilon_0 \sim N(0, \sigma^2),$$

若令 $v = \ln y, u = \ln x, c = \ln a, \varepsilon = \ln \varepsilon_0$, 则 $v = c + bu + \varepsilon, \varepsilon \sim N(0, \sigma^2)$.

$$\textcircled{5} y = a e^{bx + \varepsilon}, \varepsilon \sim N(0, \sigma^2),$$

若令 $v = \ln y, c = \ln a$, 则 $v = c + bx + \varepsilon, \varepsilon \sim N(0, \sigma^2)$.

$$\textcircled{6} y = \frac{1}{a + b e^{-x} + \varepsilon}, \varepsilon \sim N(0, \sigma^2),$$

若令 $v = \frac{1}{y}, u = e^{-x}$, 则 $v = a + bu + \varepsilon, \varepsilon \sim N(0, \sigma^2)$.

9.1.2 典型题型解析

1. 一元线性回归模型的辨识

在解答这类问题时, 要抓住一元线性回归模型式中各参数的含义, 特别是 ε 是服从正态分布 $N(0, \sigma^2)$ 这一特点即可.

例 9.1.1 在一元线性回归数学模型 $Y = a + bx + \varepsilon, \varepsilon \sim N(0, \sigma^2)$ 中, 下面说法成立的是 ().

A. x 是随机变量, y, ε 是确定性变量 B. y 是随机变量, x, ε 是确定性变量

C. ε 是随机变量, x, y 是确定性变量 D. y, ε 是随机变量, x 是确定性变量

解 分析一元线性回归数学模型的特征是 y 和 ε 是随机变量, x 为常规变量(确定性变量), 只要抓住这点, 则容易得出正确结论.

答案 D.

例 9.1.2 在一元线性回归数学模型 $Y = a + bx + \varepsilon$ 中, 下面说法正确的是 ().

A. ε 是确定性变量 B. ε 是服从 Poisson 分布的随机变量

C. ε 是服从正态分布的随机变量 D. ε 是介于 (0, 1) 之间的小数

解 根据一元线性回归数学模型的定义, ε 是服从正态分布 $N(0, \sigma^2)$ 的随机变量, 故选 C.

2. 回归方程概念及回归系数确定方法辨识

回归方程的相关问题, 一般在掌握基本概念的基础上比较容易解答.

例 9.1.3 一元线性回归的数学模型 $Y = a + bx + \varepsilon, \varepsilon \sim N(0, \sigma^2)$, 回归方程为 _____, $\hat{a} =$ _____ 称为 _____, $\hat{b} =$ _____ 称为 _____.

解 回归方程为 $\hat{y} = \hat{a} + \hat{b}x$, $\hat{a} = \bar{y} - \bar{x}\hat{b}$ 称为回归常数, $\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$,

称为回归系数.

例 9.1.4 一元线性回归数学模型 $Y = a + bx + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ 中, 回归方程为().

A. $y = a + bx$

B. $\hat{y} = \hat{a} + \hat{b}x$

C. $y = bx + \varepsilon$

D. $y = a + bx + \varepsilon$

解 根据回归方程的定义, $\hat{y} = \hat{a} + \hat{b}x$ 为回归方程, 故选 B.

3. 求解回归方程

若问题是如有一些样本值, 其值之间存在较明显的线性相关, 则求回归方程较易, 只需根据公式计算 $\hat{a}$ 和 $\hat{b}$ 即可, 但计算量大而繁复, 故选择一个表格化、合理的解题过程较好些. 另一类问题是回归方程的某些参数已知, 而求另外一些参数或其他相关问题, 这类问题只需掌握参数之间的关系, 一般能求出.

例 9.1.5 炼钢基本上是氧化脱碳的过程, 钢液原来的含碳量的多少直接影响到冶炼时间的长短, 下表是某平炉 34 炉的熔毕碳 (即全部炉料熔化完毕时钢液的含碳量) 与精炼时间 (从熔毕至出钢, 冶炼所需的时间) 的生产记录, 试求熔毕碳 x 与精炼时间 Y 的回归方程.

编号	熔毕碳 x (0.01%)	精炼时间 Y (分)	编号	熔毕碳 x (0.01%)	精炼时间 Y (分)
1	180	200	18	116	100
2	104	100	19	123	110
3	134	135	20	151	180
4	141	125	21	110	130
5	204	235	22	108	110
6	150	170	23	158	130
7	121	125	24	107	115
8	151	135	25	180	240
9	147	155	26	127	135
10	145	165	27	115	120
11	141	135	28	191	205
12	144	160	29	190	220
13	190	190	30	153	145
14	190	210	31	155	160
15	161	145	32	177	185
16	165	195	33	177	205
17	154	150	34	143	160

解 根据回归方程的形式 $\hat{y} = \hat{a} + \hat{b}x$, 需先计算 $\hat{a}$, $\hat{b}$.

$$\bar{x} = 150.09, \bar{y} = 158.23.$$

$$S_{xx} = \sum_{i=1}^{34} (x_i - \bar{x})^2 = \sum_{i=1}^{34} x_i^2 - \frac{1}{34} \left(\sum_{i=1}^{34} x_i \right)^2 = 25\,462.7,$$

$$S_{yy} = \sum_{i=1}^{34} (y_i - \bar{y})^2 = \sum_{i=1}^{34} y_i^2 - \frac{1}{34} \left(\sum_{i=1}^{34} y_i \right)^2 = 50\,094.0,$$

$$S_{xy} = \sum_{i=1}^{34} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{34} x_i y_i - \frac{1}{34} \left(\sum_{i=1}^{34} x_i \right) \left(\sum_{i=1}^{34} y_i \right) = 32\,325.3.$$

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{32\,325.3}{25\,462.7} = 1.27,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 158.23 - 1.27 \times 150.09 = -32.38,$$

所以, 回归方程为 $\hat{y} = -32.38 + 1.27x$.

4. 求解回归方程, 并做预测

在一元线性回归模型中, 当 $x = x_0$ 处, 通过回归方程 $\hat{y} = \hat{a} + \hat{b}x$ 可预测 Y 的取值 $Y_0 = \hat{a} + \hat{b}x_0$, 也可以求出置信水平为 $1 - \alpha$ 的预测区间

$$\left(\hat{a} + \hat{b}x_0 \pm t_{\frac{\alpha}{2}}(n-2)\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right).$$

例 9.1.6 某厂的耗电量(换算成金额)与生产量(换算成金额)的统计数据如下表, 求出回归方程; 若要求下月的生产量为 840, 问耗电量估计为多少?

月份	生产量 x	耗电量 y	月份	生产量 x	耗电量 y
1	511	308	7	745	445
2	546	332	8	752	458
3	587	352	9	780	475
4	609	371	10	806	490
5	642	386	11	831	498
6	723	436	12	850	510

解 先求回归方程, 由表上数据得

$$\bar{x} = \frac{8\ 382}{12} = 698.5, \quad \bar{y} = \frac{5\ 061}{12} = 421.75.$$

$$S_{xx} = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = 6\ 001\ 206 - 12 \times 698.5^2 = 146\ 379,$$

$$S_{yy} = \sum_{i=1}^n y_i^2 - ny^2 = 2\ 187\ 343 - 12 \times 421.75^2 = 52\ 866.3,$$

$$S_{xy} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = 3\ 622\ 982 - 12 \times 698.5 \times 421.75 = 87\ 873.5.$$

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{87\ 873.5}{146\ 379} = 0.600,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 421.75 - 0.600 \times 698.5 = 2.650.$$

得回归方程为 $y = 2.650 + 0.600x$.

生产量为 840 时, 耗电量的预测运算如下:

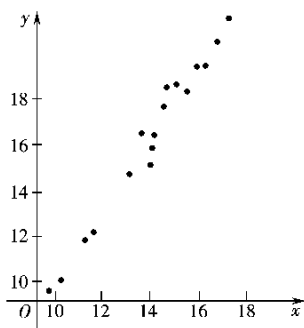
$$y = 2.650 + 0.600 \times 840 = 506.65.$$

例 9.1.7 下表列出了 18 个 5~8 岁儿童的重量(这是容易测得的)和体积(这是难以测量的):

重量 x (kg)	17.1	10.5	13.8	15.7	11.9	10.4	15.0	16.0	17.8	15.8
体积 y (dm ³)	16.7	10.4	13.5	15.7	11.6	10.2	14.5	15.8	17.6	15.2
重量 x (kg)	15.1	12.1	18.4	17.1	16.7	16.5	15.1	15.1		
体积 y (dm ³)	14.8	11.9	18.3	16.7	16.6	15.9	15.1	14.5		

(1)画出散点图;(2)求 Y 关于 x 的线性回归方程 $\hat{y} = \hat{a} + \hat{b}x$;(3)求 $x = 14.0$ 时 Y 的置信度为 0.95 的预测区间.

解 (1)散点图如下



(2) $\bar{x} = 15.005\ 555\ 56$, $\bar{y} = 14.722\ 222\ 22$, $n = 18$,

$$\sum_{i=1}^{18} x_i = 270.1, \sum_{i=1}^{18} y_i = 265, \sum_{i=1}^{18} y_i^2 = 3\ 996.14,$$

$$\sum_{i=1}^{18} x_i^2 = 4\ 149.39, \sum_{i=1}^{18} x_i y_i = 4\ 071.71,$$

$$S_{xx} = \sum_{i=1}^{18} x_i^2 - \frac{1}{18} \left(\sum_{i=1}^{18} x_i \right)^2 = 96.389\ 444\ 46,$$

$$S_{yy} = \sum_{i=1}^{18} y_i^2 - \frac{1}{18} \left(\sum_{i=1}^{18} y_i \right)^2 = 94.751\ 111\ 12,$$

$$S_{xy} = \sum_{i=1}^{18} x_i y_i - \frac{1}{18} \left(\sum_{i=1}^{18} x_i \right) \left(\sum_{i=1}^{18} y_i \right) = 95.237\ 777\ 78,$$

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = 0.988051942 \doteq 0.988\ 0,$$

$$\hat{a} = -0.104\ 046\ 085 \doteq -0.104\ 0.$$

故线性回归方程为

$$\hat{y} = -0.104 + 0.988x.$$

(3) $1 - \alpha = 0.95$, $\alpha = 0.05$, $t_{0.025}(16) = 2.119\ 9$, $Q_e = 0.651\ 239\ 832$.

$$\hat{\sigma}^2 = \frac{Q_e}{16} = (0.201748579)^2.$$

$$t_{\frac{\alpha}{2}}(16) \hat{\sigma} \sqrt{1 + \frac{1}{18} + (14 - \bar{x})^2 / S_{xx}} = 0.442.$$

在 $x = 14$ 处 Y 的观察值的预测值为

$$\hat{y}|_{x=14} = -0.104 + 0.988x|_{x=14} = 13.728.$$

在 $x = 14$ 处观察值 Y 的置信水平为 0.95 的预测区间为

$$(13.728 \pm 0.442) = (13.29, 14.17).$$

5. 回归方程显著性检验

检验回归方程是否有明显的线性关系, 从而得出回归方程的价值. 其主要的方法有 t 检验、 F 检验和相关系数检验. 不同方法得出的结论是一致的. 解题时只需按公式和方法去做即可.

例 9.1.8 在一元线性回归中, 若相关系数 $r > 0$ 则().

- A. y 与 x 之间存在着密切的线性关系
- B. y 与 x 之间不存在密切的线性关系
- C. y 与 x 无线性关系
- D. 回归方程单调增加

解 由于 $r > 0$ 取正值, 此时 $\hat{b} > 0$, 称 x 与 y 正相关. 直线的斜率为正值, 回归方程单调增加, 故应选 D.

例 9.1.9 对例 9.1.6 所给的数据, 考察生产量(x)与耗电量(y)之间有无显著的线性关系(显著性水平 $\alpha = 0.05$).

$$\text{解 } S_R = \hat{b}S_{xy} = 0.600 \times 87\,873.5 = 52\,724.1,$$

$$S_E = S_{yy} - S_R = 52\,866.3 - 52\,724.1 = 142.2.$$

$$\text{检验统计量 } F = \frac{S_R(n-2)}{S_E} = \frac{52\,724.1 \times 10}{142.2} = 3\,708 > F_{0.05}(1, 10) = 4.96.$$

故 x 与 y 之间具有显著的线性相关关系.

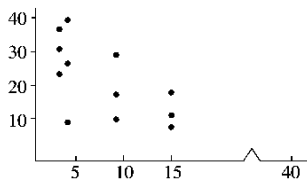
6. 求可化为线性回归的回归曲线方程

当问题明确了变量之间的数量关系类型, 则按所给方法进行即可; 当没有明确变量之间关系类型, 可通过画出散点图进行观察, 进而估计出变量之间的关系类型. 在此基础上按所给方法计算.

例 9.1.10 槲寄生是一种寄生在大树上部树枝上的寄生植物, 它喜欢寄生在年轻的大树上. 下面给出在一定条件下完成的试验中采集的数据. ①做出 (x_i, y_i) 的散点图; ②令 $z_i = \ln y_i$, 做出 (x_i, z_i) 的散点图; ③以模型 $Y = ae^{bx}\epsilon$, $\ln \epsilon \sim N(0, \sigma^2)$ 拟合数据, 其中 a, b, σ^2 与 x 无关. 试求曲线回归方程 $\hat{y} = \hat{a} \exp(\hat{b}x)$.

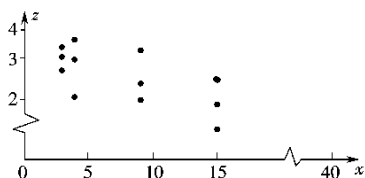
x : 大树的年龄(年)	3	4	9	15	40
y : 每株大树上槲寄生的株数	28	10	15	6	1
	33	36	22	14	1
	22	24	10	9	

解 (1)散点图如下



(2)令 $z_i = \ln y_i$, 得数据如下表

z_i \ x_i	3	4	9	15	40
	3.33	2.30	2.71	1.79	0
	3.50	3.58	3.09	2.64	0
	3.09	3.18	2.30	2.20	



(3)将 $Y = ae^{bx}\epsilon$ 取对数, 得

$$\ln Y = \ln a + bx + \ln \epsilon.$$

令 $z = \ln Y$ 则回归模型为

$$z = \ln a + bx + \ln \epsilon,$$

其中 $\ln \epsilon \sim N(0, \sigma^2)$.

$$n = 14, \sum_{i=1}^{14} x_i = 173, \sum_{i=1}^{14} x_i^2 = 4193, \sum_{i=1}^{14} z_i = 33.71,$$

$$S_{xx} = 2055.214286, \sum_{i=1}^{14} x_i z_i = 238.35,$$

$$S_{xz} = -178.2525455, \hat{b} = -0.086731854,$$

$$\ln \hat{a} = 3.47987445,$$

$$\exp(\ln \hat{a}) = 32.45564701.$$

得回归曲线方程为

$$\hat{y} = 32.4556e^{-0.0867318x}.$$

7. 其他

例 9.1.11 在一元线性回归中, 离差平方和为 $S_{yy} =$ _____, 回归误差平方和为 $S_R =$ _____, 残差平方和 $S_E =$ _____, S_{yy} , S_R , S_E 满足的关系式为 _____.

解 $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2,$

$$S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2,$$

$$S_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

它们之间的关系为 $S_{yy} = S_R + S_E$.

例 9.1.12 一元线性回归的正规方程

$$\begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

有唯一解的充要条件是 $x_1, x_2, \dots, x_n$ 不全相等.

证明 由线性代数知识知, 有唯一解的充要条件是

$$D = \begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} \neq 0$$

而 $D = n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 = n \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 \right]$

$$= n \left[\sum_{i=1}^n x_i^2 - n \frac{(\sum_{i=1}^n x_i)^2}{n^2} \right] = n \left[\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right]$$

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

正规方程有唯一解的充要条件是 $\sum_{i=1}^n (x_i - \bar{x})^2 \neq 0$. 亦即存在某个 i 使 $x_i - \bar{x} \neq 0, i = 1, 2, \dots, n$ 故 $x_1, x_2, \dots, x_n$ 不全相等.

同步练习 9.1

1. 在一元线性回归分析中, r 称为相关系数. 若 $r = 1$ 则称 x 与 y _____ 若 $r = -1$ 则称 x 与 y _____ 若 $r = 0$ 则称 x 与 y _____.

2. 在一元线性回归中, 若相关系数 $r = 1$ 则().

A. y 与 x 之间满足线性关系 $y = a + bx$ 并且 $b < 0$

B. y 与 x 之间满足线性关系 $y = a + bx$ 并且 $b > 0$

C. y 与 x 之间不具有线性关系 并且 x 为可控变量

D. y 与 x 之间不具有线性关系 y 是可控变量

3. 在一元线性回归中, 对任一组观察值 $(x_i, y_i), i = 1, 2, \dots, n$, 总可以得到一个线性回归方程 $y = \hat{a} + \hat{b}x$, 从而对于给定的 x_0 值, 总可以得到 y 的预测值 y_0 , 试分析上述说法是否正确.

4. 下表数据是退火温度 $x(^{\circ}\text{C})$ 对黄铜延性 Y 效应的试验结果, Y 是以延长度计算的.

$x(^{\circ}\text{C})$	300	400	500	600	700	800
$y(\%)$	40	50	55	60	67	70

画出散点图并求 Y 对 x 的线性回归方程.

5. 假设儿子的身高(y)与父亲的身长(x)适合一元正态线性回归模型, 观察了 10 对英国父子的身长(英寸)如下:

x	60	62	64	65	66	67	68	70	72	74
y	63.6	65.2	66	65.5	66.9	67.1	67.4	63.3	70.1	70

(1) 建立 y 关于 x 的回归方程;

(2) 对线性回归方程作假设检验(检验水平取为 0.05);

(3) 给出 $x_0 = 69$ 时, Y_0 的置信度为 95% 的预测区间.

6. 由生产线上 20 名工人组成的一个随机样本的年龄和有效定额由下表给出:

年龄	定额	年龄	定额	年龄	定额	年龄	定额
44	61	52	67	51	60	34	70
44	41	43	73	50	78	51	60
45	89	47	94	61	78	48	67
43	76	54	96	47	74	51	72
40	79	43	77	62	82	57	80

(1) 试建立年龄和有效定额之间线性关系的回归方程;

(2) 求相关系数;

(3) 检验两个变量间是否存在线性相关关系($\alpha = 0.05$).

7. 某地区车祸次数与机动车拥有量的 11 年统计数据如下表:

年度	车祸次数 Y /千次	汽车拥有量 x /万辆	年度	车祸次数 Y /千次	汽车拥有量 x /万辆
1	166	352	7	227	529
2	153	373	8	238	577
3	177	411	9	268	641
4	201	441	10	268	692
5	216	462	11	274	742
6	208	490			

(1)建立车祸次数对汽车拥有量的回归方程；

(2)假设拥有 1000 万辆汽车,求车祸次数 0.95 的预测区间.

同步练习 9.1 参考答案

1. 正相关 ;负相关 ;零相关(即无线性相关关系)

2. B

3. 不正确. 由于对不具有线性关系的一组观察值 (x_i, y_i) , $i = 1, 2, \dots, n$, 用最小二乘估计法, 总可以建立线性回归方程 $y = \hat{a} + \hat{b}x$, 因此, 必须进行线性相关性检验才能确定 y 与 x 是否具有线性相关关系. 并且, 当 y 与 x 不具有线性相关关系时, 所得的回归方程 $y = \hat{a} + \hat{b}x$ 将毫无意义.

4. $\hat{y} = 24.6287 + 0.05886x$.

5. (1) $\hat{y} = 41.7072 + 0.3713x$.

(2)线性相关关系显著.

(3) $x_0 = 69$ 时, Y_0 置信度为 0.95 的预测区间的观察值为 (63.0432, 71.6106).

6. (1) $\hat{y} = 54.6123 + 0.3948x$.

(2) $r = 0.215$.

(3) $t = 0.934$ 两变量间不存在线性相关关系.

7. (1) $\hat{y} = 55.85 + 0.3120x$.

(2) (336.2, 399.4).

9.2 多元线性回归

9.2.1 知识要点

1. 多元线性回归模型

设随机变量 Y 与 p 个变量 $x_1, x_2, \dots, x_p$ 有关系式

$$\left. \begin{aligned} Y &= b_0 + b_1 x_1 + \dots + b_p x_p + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2), \end{aligned} \right\}$$

其中 $b_0, b_1, \dots, b_p, \sigma^2$ 为未知参数, 称为 p 元线性回归模型.

此时有: $Y \sim N(b_0 + b_1 x_1 + \dots + b_p x_p, \sigma^2)$.

2. 回归系数 $b_0, b_1, \dots, b_p$ 估计法

设样本值为

$(x_{11} \ x_{12} \ \cdots \ x_{1p} \ y_1) \ (x_{21} \ x_{22} \ \cdots \ x_{2p} \ y_2) \ \cdots \ (x_{n1} \ x_{n2} \ \cdots \ x_{np} \ y_n)$,
采用极大似然估计法估计参数 $\hat{\beta}_0 \ \hat{\beta}_1 \ \cdots \ \hat{\beta}_p$ 求法如下.

$$\text{记 } \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ 则 } \hat{\mathbf{B}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

3. 回归方程显著性检验

多元回归方程显著性检验与一元线性回归类似.

$$\text{检验统计量 } F = \frac{S_R / p}{S_E / (n - p - 1)} \sim F(p, n - p - 1),$$

在给定的显著性水平 α 下, 拒绝域为 $F \geq F_{\alpha}(p, n - p - 1)$.

9.2.2 典型题型解析

1. 求多元回归方程

例 9.2.1 某种化工产品的得率 Y 与反应温度 x_1 、反应时间 x_2 及某反应物浓度 x_3 有关. 今得试验结果如下表所示, 其中 $x_1 \ x_2 \ x_3$ 均为二水平且均以编码形式表达.

x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
得率	7.6	10.3	9.2	10.2	8.4	11.1	9.8	12.6

(1) 设 $\mu(x_1 \ x_2 \ x_3) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$, 求 Y 的多元线性回归方程;

(2) 若认为反应时间不影响得率, 即认为 $\mu(x_1 \ x_2 \ x_3) = \beta_0 + \beta_1 x_1 + \beta_3 x_3$, 求 Y 的多元线性回归方程.

解 (1) 矩阵 $\mathbf{X}, \mathbf{Y}, \mathbf{B}$ 分别为

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} 7.6 \\ 10.3 \\ 9.2 \\ 10.2 \\ 8.4 \\ 11.1 \\ 9.8 \\ 12.6 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix},$$

则所要求的线性回归模型为

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \varepsilon, \varepsilon \sim N(0, \sigma^2).$$

其正规方程为

$$\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{X}^T \mathbf{Y},$$

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}, \quad \mathbf{X}^T \mathbf{Y} = \begin{bmatrix} 79.2 \\ 4.6 \\ 4.4 \\ 9.2 \end{bmatrix},$$

故

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{8} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & \frac{1}{8} \end{bmatrix},$$

所以

$$\hat{\mathbf{B}} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{bmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \begin{bmatrix} 9.9 \\ 0.575 \\ 0.55 \\ 1.15 \end{bmatrix}.$$

所以多元回归方程为

$$\hat{y} = 9.9 + 0.575x_1 + 0.55x_2 + 1.15x_3.$$

(2)若认为 $\mu(x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_3 x_3$, 则引入的 8×3 矩阵就是上述 $\mathbf{X}$ 中删去第 3 列后所得的矩阵, 即

$$\mathbf{M} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_3 \end{bmatrix}.$$

模型 $y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ 的正规方程为

$$\mathbf{M}^T \mathbf{M} \boldsymbol{\beta} = \mathbf{M}^T \mathbf{Y}.$$

$$\mathbf{M}^T \mathbf{M} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \quad \mathbf{M}^T \mathbf{Y} = \begin{bmatrix} 79.2 \\ 4.6 \\ 9.2 \end{bmatrix},$$

故

$$(\mathbf{M}^T \mathbf{M})^{-1} = \begin{bmatrix} \frac{1}{8} & 0 & 0 \\ 0 & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{8} \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 9.9 \\ 0.575 \\ 1.15 \end{bmatrix},$$

得多元回归方程为 $\hat{y} = 9.9 + 0.575x_1 + 1.15x_3$.

2. 线性回归的显著性检验

多元线性回归的显著性检验与一元线性回归显著性检验原理一致, 但使用的公式有所不

同,只要对公式认识正确,正确使用即可.

例 9.2.2 某市自来水厂欲对该市的用水量进行预测,经调查分析,认为城市的用水量与人口总数和工业总产值有关,现将该市连续 10 年的有关统计资料(用水量)列表如下(用水量以百万 m^3 为单位;人口以万人为单位;工业总产值以亿元为单位):

年度	1	2	3	4	5	6	7	8	9	10
用水量 y	176	180	188	200	205	224	226	239	256	270
人口 x_1	105	108	110	112	115	116	120	121	122	124
工业总产值 x_2	78	80	83	91	100	100	115	126	128	135

试用这些数据求出回归方程,并在显著性水平 $\alpha=0.01$ 下,检验线性相关性.

解 这是一个二元线性回归问题,计算可列表进行.

序号	y	x_1	x_2	y^2	x_1^2	x_2^2	$x_1 y$	$x_2 y$	$x_1 x_2$
1	176	105	78	30 976	11 025	6 084	18 480	13 728	8 190
2	180	108	80	32 400	11 664	6 400	19 440	14 400	8 640
3	188	110	83	35 344	12 100	6 889	20 680	15 604	9 130
4	200	112	91	40 000	12 544	8 281	22 400	18 200	10 192
5	205	115	100	42 025	13 225	10 000	23 575	20 500	11 500
6	224	116	100	50 176	13 456	10 000	25 980	22 400	11 600
7	226	120	115	51 076	14 400	13 225	27 120	25 990	13 800
8	239	121	126	57 121	14 641	15 876	28 919	30 114	15 246
9	256	122	128	65 536	14 884	16 384	31 232	32 768	15 616
10	270	124	135	72 900	15 376	18 225	33 480	36 450	16 740
Σ	2 164	1 153	1 036	477 554	133 315	111 364	251 310	230 154	120 654

由表中数据可得

$$\bar{y} = \frac{1}{10} \times 2\,164 = 216.4,$$

$$\bar{x}_1 = \frac{1}{10} \times 1\,153 = 115.3,$$

$$\bar{x}_2 = \frac{1}{10} \times 1\,036 = 103.6.$$

$$l_{11} = 133\,315 - 10 \times 115.3^2 = 374.1,$$

$$l_{22} = 111\,364 - 10 \times 103.6^2 = 4\,034.4,$$

$$l_{12} = 120\,654 - 10 \times 115.3 \times 103.6 = 1\,203.2,$$

$$l_{1y} = 251\,310 - 10 \times 115.3 \times 216.4 = 1\,800.8,$$

$$l_{2y} = 230\,154 - 10 \times 103.6 \times 216.4 = 5\,963.6.$$

正规方程为

$$\begin{cases} 374.1b_1 + 1\,203.2b_2 = 1\,800.8, \\ 1\,203.2b_1 + 4\,034.4b_2 = 5\,963.6. \end{cases}$$

解方程组,得

$$\hat{b}_1 = 1.457\,4, \hat{b}_2 = 1.043\,5,$$

故
$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}_1 - \hat{b}_2 \bar{x}_2$$

$$= 216.4 - 1.4574 \times 115.3 - 1.0435 \times 103.6 \\ = -59.745.$$

所以用水量与人口和工业总产值的回归方程为

$$y = -59.745 + 1.4574x_1 + 1.0435x_2.$$

下面进行线性相关检验($\alpha = 0.01$).

零假设为 $H_0: b_1 = b_2 = 0$.

$$\text{计算得 } L_{yy} = \sum_{i=1}^n y_i^2 - n\bar{y}^2 \\ = 477\,554 - 10 \times 216.4^2 \\ = 9\,264.4,$$

$$U = b_1 L_{1y} + b_2 L_{2y} \\ = 1.4574 \times 1\,800.8 + 1.0435 \times 5\,963.6 \\ = 8\,847.5,$$

$$Q = L_{yy} - U \\ = 9\,264.4 - 8\,847.5 \\ = 416.9.$$

$$\text{于是 } F = \frac{\frac{U}{K}}{\frac{Q}{n-K-1}} = \frac{\frac{8\,847.5}{2}}{\frac{416.9}{10-2-1}} = 74.28.$$

查表得 $F_{0.01}(2, 7) = 9.55$.

由 $F > F_{0.01}(2, 7)$ 故在 $\alpha = 0.01$ 下, 拒绝零假设 H_0 , 即认为 y 与 x_1, x_2 的线性相关关系是显著的.

同步练习 9.2

1. 在某项钢材的新型规范试验中, 研究含碳量(x_1)和回火温度(x_2)对它的伸长率(Y)的关系. 15 批生产试样结果如下:

含碳量 x_1	57	64	69	58	58	58	58	58	58	57
回火温度 x_2	535	535	535	460	460	460	490	490	490	460
伸长率 y	19.25	17.5	18.25	16.25	17.00	16.75	17.00	16.75	17.25	16.75
含碳量 x_1	64	69	59	64	69					
回火温度 x_2	435	460	490	467	490					
伸长率 y	14.75	12.00	17.75	15.50	15.50					

根据经验, Y 关于 x_1, x_2 有二元线性回归关系

$$Y = b_0 + b_1x + b_2x_2 + \varepsilon, \varepsilon \sim N(0, \sigma^2).$$

(1) 求出二元线性回归方程;

(2) 检验线性回归是否显著($\alpha = 0.05$);

(3) 当 $x_1 = 70, x_2 = 540$ 时对 Y 做区间估计($1 - \alpha = 0.95$).

2. 研究同一地区土壤所含植物可给态磷的情况得 18 组数据如下表所示:

土壤子样	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
x_1	0.4	0.4	3.1	0.6	4.7	1.7	9.4	10.1	11.6	12.6	10.9	23.1	23.1	21.6	23.1	1.9	26.8	29.9
x_2	53	23	19	34	24	65	44	31	29	58	37	46	50	44	56	36	58	51
x_3	158	163	37	157	59	123	46	117	173	112	111	114	134	73	168	143	202	124
y	64	60	71	61	54	77	81	93	93	51	76	96	77	93	95	54	168	99

其中 x_1 : 土壤内所含无机磷浓度 ;

x_2 : 土壤内溶于 K_2CO_3 溶液并受溴化物水解的有机磷浓度 ;

x_3 : 土壤内溶于 K_2CO_3 溶液但不受溴化物水解的有机磷浓度 ;

Y 种在 $20^\circ C$ 土壤内的玉米中的可给态磷.

已知 Y 对 x_1, x_2, x_3 存在线性回归关系, 试求出经验回归方程, 并检验线性回归是否显著 ($\alpha = 0.05$).

3. 一种合金在某种添加剂的不同浓度之下, 各做三次试验, 数据如下:

浓度 x	10.0	15.0	20.0	25.0	30.0
抗压强度 y	25.2	29.8	31.2	31.7	29.4
	27.3	31.1	32.6	30.1	30.8
	28.7	27.8	29.7	32.3	32.8

(1) 作散点图, (2) 以模型 $Y = b_0 + b_1x + b_2x^2 + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ 的拟合数据, 其中 b_0, b_1, b_2, σ^2 与 x 无关, 求回归方程 $\hat{y} = \hat{b}_0 + \hat{b}_1x + \hat{b}_2x^2$.

同步练习 9.2 参考答案

1. (1) $\hat{y} = 10.514 - 0.216x_1 + 0.040x_2$.

(2) 线性回归显著.

(3) (14.697, 19.138).

2. $\hat{y} = 43.65 + 1.78x_1 - 0.08x_2 + 0.16x_3$, 线性回归显著.

3. $\hat{y} = 19.0333 + 1.0086x - 0.020381x^2$.

疑难解析

1. 如何区别所讨论的问题是回归分析还是方差分析?

答 回归分析与方差分析都是考察某一指标与试验因素(条件)的关系. 回归分析考察的是因素取值与指标取值存在一种什么样的相关关系, 而方差分析所考察的是因素对指标的影响是否显著.

因素可以划分成两大类: 一类是属性的, 一类是数量的. 属性因素一般无数量大小可言, 只是性质的不同、机器的型号、材料的品质、加工工艺等; 数量因素可以是一定范围内的取值, 如产品的合格率、产量、试验的温度、人的身高、体重等; 也有些因素本来是数量而非属性, 但可属性化, 如施肥量是某个数量, 但有时将它局限在某些范围内而分为高中低几个层次, 从而属性化.

当所研究问题的因素是属性的时候, 问题归为方差分析范畴; 当所考虑的因素是数量的时候, 则问题属于回归分析的范围.

2. 进行回归分析需满足的基本条件是什么?

答 在进行回归分析过程中,对参数性检验对象要求必须满足以下三个基本条件.①正态性.即被检验的对象或者因变量必须是服从正态分布 $N(\mu, \sigma^2)$ 的随机变量.②方差齐性.被检验的各个总体的方差应该是相等的.③独立性.对被检验的各对观察数据而言,从概率意义上应理解为是独立取得的.

处理实际问题时常常不能事先预知这三个条件是否满足.关于正态性可由大数定律与中心极限定理来确定,方差齐性可用 F 检验施行,独立性一般凭实际经验判断.一般情况有一个大致的结论即可进行回归分析.

3. 回归分析与相关分析的区别与联系是什么?两种分析能否互相取代?

答 相关关系是一种不确定性关系,也可以分为自变量与因变量加以考察.因变量一般取可以测量的随机变量,而自变量在许多情况下不是随机的,是可控制的普通变量.它表现为因变量的取值随自变量的变化而呈现一定的统计规律性,即当自变量取值确定时,因变量的取值也有确定的概率分布与之对应.

相关分析一般是研究随机变量与随机变量之间的相关关系的,而回归分析研究随机变量与非随机变量之间的相关关系.两者所使用的概念、理论和方法有所不同,得到的结果含义也不相同,但结果的形式却几乎完全一致,所以,从应用与计算角度上看,两者没有必要加以严格区别.而由于回归分析在数学处理上更为简便,因而无论自变量如何,都可作为非随机的普通变量看待,而用回归分析方法研究变量间的相关关系.因此,从这个意义上说,在某种场合下两种分析可互相取代.

4. 试分析回归系数的最小二乘估计与极大似然估计的异同.

答 现以一元正态线性回归为例分析.其数学模型为

$$\begin{cases} Y = a + bx + \varepsilon, \\ \varepsilon \sim N(0, \sigma^2). \end{cases}$$

最小二乘估计是对 x, y 的 n 对试验值 $(x_i, y_i) (i=1, 2, \dots, n)$ 做离差平方和

$$Q = \sum_{i=1}^n (y_i - a - bx_i)^2,$$

利用微积分中的极值方法,求出使 $Q(\hat{a}, \hat{b}) = \min Q(a, b)$ 的 $\hat{a}$ 和 $\hat{b}$, 则 $\hat{y} = \hat{a} + \hat{b}x$ 为所求线性回归方程.

极大似然估计法是由 $y_i = a + bx_i + \varepsilon, \varepsilon \sim N(0, \sigma^2), y_i \sim N(a + bx_i, \sigma^2)$, 得极大似然函数

$$L = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - a - bx_i)^2 \right].$$

要使 L 取最大值, 则应使 $\sum_{i=1}^n (y_i - a - bx_i)^2$ 为最小, 用与最小二乘法中同样的极值方法, 求得使

$$Q = \sum_{i=1}^n (y_i - a - bx_i)^2$$

最小的 $\hat{a}$ 和 $\hat{b}$, 得线性回归方程 $\hat{y} = \hat{a} + \hat{b}x$.

由于两种方法讨论的都是 $Q = \sum_{i=1}^n (y_i - a - bx_i)^2$, 且由此求得 $\hat{a}$ 和 $\hat{b}$, 故最小二乘法与极

大似然法的结果是相同的,只是求解的出发点不同罢了.

5. 利用线性回归预测时,影响预测精度的主要因素是什么?

答 回归方程 $\hat{y} = \hat{a} + \hat{b}x$ 已经确定为检验确认回归显著的前提下,则对给定的 x_0, y_0 ,置信度为 $1 - \alpha$ 的预测区间为

$$\left(\hat{y}_0 - t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{y}_0 + t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right),$$

其中 $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$.

由此可知影响预测精度的主要因素为:① σ^2 ,一般 σ^2 越小,精度越高;② n , n 越大,精度越高,所以应尽量扩大样本容量;③自变量的取值 x_i 应尽量避免过于集中,但预测点 x_0 离 $\bar{x}$ 越近时精度越高.

6. 如何将可化为线性回归的非线性回归问题线性化?

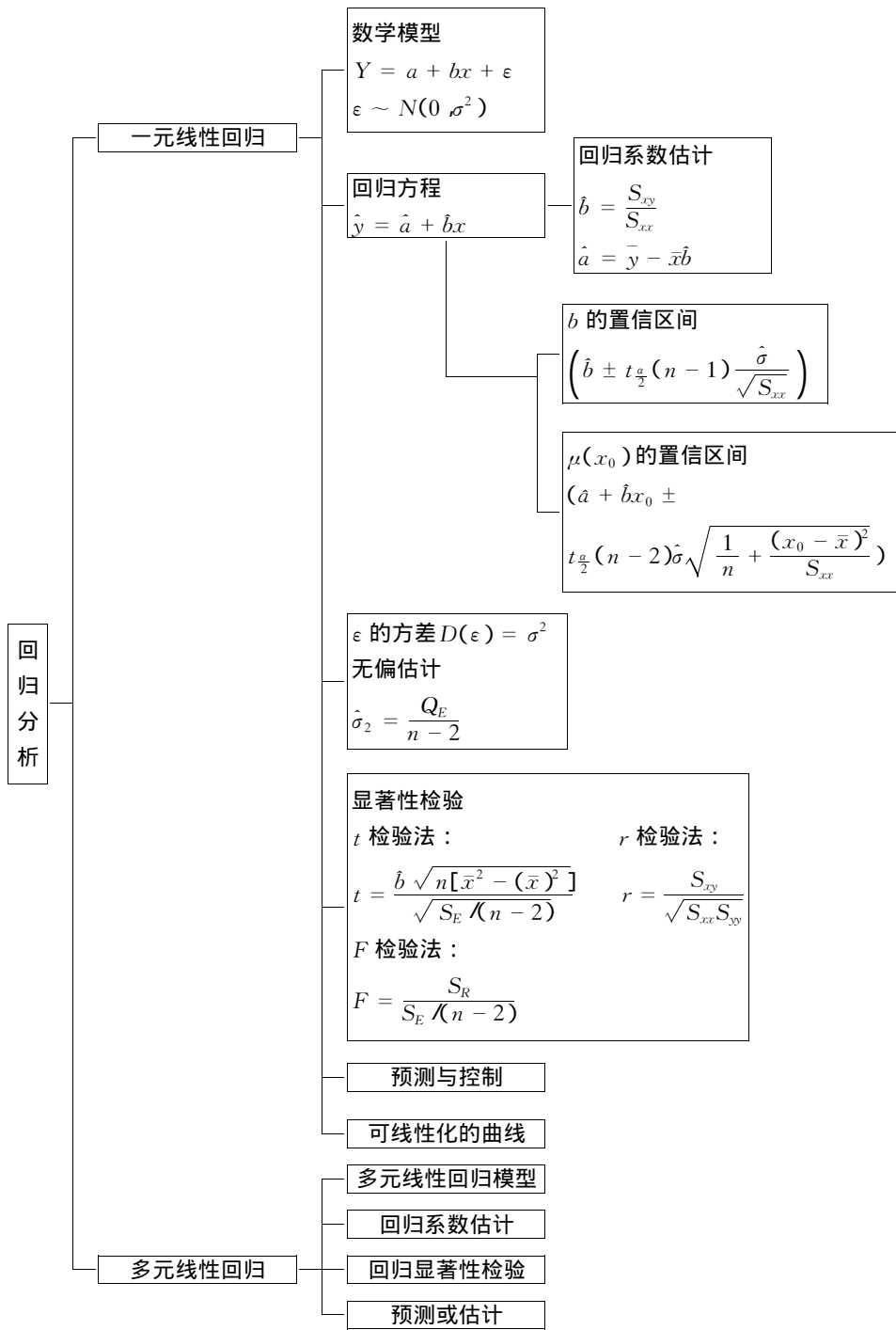
答 当具有相关关系的两个变量不具有线性相关关系,而具有某种曲线相关关系时,可以通过适当的变量代换,将变量间的关系化为线性的形式,即通过变量转换对其做线性化处理. 在两个变量回归的条件下,一般常用的有倒数变换与对数变换,使曲线问题化为变换后的直线问题. 所处理的自变量与因变量间的关系,可以是双曲函数、幂函数、指数函数、对数函数等.

其中正确选择曲线类型,是正确地进行变量转换的前提,而正确的转换关系又是提高曲线回归精确度的根本. 由散点图的形状选择线性化转换,往往不是一次就能选准,不妨同时做几种曲线加以比较,择优取之.

本章常考题型归纳

考点	常见题型	例题
一元线性回归	一元线性回归模型辨识	例 9.1.1, 例 9.1.2
	求一元线性回归方程	例 9.1.5
	回归显著性检验	例 9.1.8, 例 9.1.9
	估计或预测	例 9.1.6
	求预测区间	例 9.1.7
	求可化为线性回归的回归方程	例 9.1.10
多元线性回归	多元线性回归模型的辨识	
	求回归方程(回归系数的估计)	例 9.2.1
	回归方程显著性检验	例 9.2.2

本章知识结构网络图



检测题(一)

1. 下表列出在不同质量下 6 根弹簧的长度：

质量 x (克)	5	10	15	20	25	30
长度 y (厘米)	7.25	8.12	8.95	9.90	10.9	11.8

(1)试将这 6 对观察值用点画在坐标纸上 ,直观上能否认为长度对于质量的回归是线性的 ;

(2)写出经验回归直线方程 ;

(3)试在 $x = 16$ 时做出 Y 的 95% 预测区间.

2. 维尼纶纤维的耐热水性能好坏可以用指标“缩醛化度” y 来衡量 ,这个指标越高 ,耐热水性能越好 .而甲醛浓度是影响缩醛化度的重要因素 ,在生产中常用甲醛浓度 $x(g/L)$ 去控制这一指标 ,经试验数据如下 :

甲醛浓度 $x(g/L)$	18	20	22	24	26	28	30
缩醛化度 y (克分子 %)	26.86	28.35	28.75	28.87	29.75	30.00	30.36

(1)试建立 x, y 的回归方程 ;

(2)求 σ^2 的无偏估计 ;

(3)对回归方程做显著性检验($\alpha = 0.05$).

3. 某医院用光电比色计检验尿汞时 ,尿汞含量(mg/L)与消光系数读数的结果如下 :

尿汞含量 x_i	2	4	6	8	10
消光系数 y_i	64	138	205	285	360

已知它们之间有关系式

$$Y_i = \alpha + \beta x_i + \epsilon_i, i = 1, 2, \dots, n,$$

其中 $\epsilon_i \sim N(0, \sigma^2)$,且各 ϵ_i 相互独立 ,试求 α, β 的点估计 ,并在显著水平 $\alpha = 0.05$ 下检验 β 是否为 38.

4. 某种合金 15 个样品的硬度(以布氏硬度为单位)和抗拉强度(kg/cm²)如下表 :

硬度	41	74	100	72	67	55	71	35	41	96	20	45	38	26	29
抗拉强度	27	42	62	43	37	31	45	21	23	51	15	22	19	17	17

(1)画出这些数据的散点图 ;

(2)求出样本线性回归方程 ;

(3)检验假设 $H_0: b = 0(\alpha = 0.05)$;

(4)假定一个新样品的硬度为 50 ,试以 0.959 0 的置信度估计其抗拉强度.

5. 在钢线碳含量对于电阻的效应研究中 ,得到以下的数据 :

碳含量 $x(\%)$	0.10	0.30	0.40	0.55	0.70	0.80	0.95
电阻 $y(20^\circ\text{C}$ 时 ,微欧)	15	18	19	21	22.6	23.8	26

(1)求线性回归方程 $\hat{y} = \hat{a} + \hat{b}x$;(2)求 b 的置信水平为 0.95 的置信区间 ;(3)求 $x = 0.50$ 处观察值 Y 的置信水平为 0.95 的预测区间.

检测题(一)参考答案

1. (1)可以 ;(2) $\hat{y} = 450.26 - 20.31x$;(3)(-591.33, 841.99).

2. (1) $\hat{y} = 22.6486 + 0.2643x$;(2) $\hat{\sigma}^2 = 0.1340$;(3) $F = 58.40$.

3. $\hat{\alpha} = -11.3, \hat{\beta} = 36.95$,接受 $\beta = 38$.

4. (2) $\hat{y} = 0.7834 + 0.5682x$; (3) $t = 18.148 > t_{\alpha}(13) = 2.160$, 故否定 H_0 ; (4) (22.69, 35.70).

5. (1) $\hat{y} = 13.9584 + 12.5503x$; (2) (11.8232, 13.2774); (3) (19.67, 20.80).

检测题(二)

1. 为考察某种毒药的剂量(mg/单位容量计)与老鼠死亡之间的关系, 取多组老鼠(每组 25 只)试验, 得到以下的数据:

剂量 x	4	6	8	10	12	14	16	18
死亡的老鼠数 y	1	3	6	8	14	16	20	21

(1) 画出散点图, 并求 Y 对 x 的线性回归方程: $\hat{y} = \hat{a} + \hat{b}x$;

(2) 估计当剂量为 7 时, 死亡的老鼠数;

(3) 求 ϵ 的方差 σ^2 的无偏估计;

(4) 检验假设 $H_0: b = 0, H_1: b \neq 0$, 取 $\alpha = 0.05$;

(5) 若回归效果显著, 求 b 的置信水平为 0.99 的置信区间;

(6) 求 $x = 9$ 处 $\mu(x)$ 的置信水平为 0.99 的置信区间;

(7) 求 $x = 9$ 处观察值 Y 的置信水平为 0.95 的预测区间.

2. 钢的强度和硬度都是反映钢质量的指标. 现在炼 20 炉中碳钢, 它们的抗拉强度 Y 与硬度 x 的 20 对试验值列于下表:

(1) 绘出散点图;

(2) 试求 Y 对 x 的经验回归直线方程;

(3) 求 σ^2 的估计量 $\hat{\sigma}^2$.

编号	x_i	y_i	编号	x_i	y_i	编号	x_i	y_i	编号	x_i	y_i
1	277	103	6	268	98	11	286	108	16	255	94
2	257	99.5	7	285	103.5	12	269	100	17	269	99
3	255	93	8	286	103	13	246	96.5	18	297	109
4	278	105	9	272	104	14	255	92	19	257	95.5
5	306	110	10	285	103	15	253	94	20	250	91

检测题(二)参考答案

1. (1) $\hat{y} = -5.964 + 1.554x$; (2) $\hat{y}|_{x=7} = 4.914$;

(3) $\hat{\sigma}^2 = 1.2321615$; (4) $|t| = 2.4469$, 拒绝 H_0 , 即认为回归效果显著;

(5) $(1.554 \pm .318)$; (6) (8.022 ± 1.588) ;

(7) $\left(8.022 \pm 3.7074 \times \hat{\sigma} \sqrt{1 + \frac{1}{8} + (9-11)^2 / 168}\right) = (8.022 \pm 4.411)$.

2. (2) $\hat{y} = 0.3156x + 14.7683$; (3) $\hat{\sigma}^2 = 4.8827$.

- A. $T \sim t(n-1)$ B. $T \sim t(n)$ C. $T \sim N(0, 1)$ D. $T \sim F(1, n)$

8. 在下列函数中, 可以作为某个随机变量分布函数的是().

- A. $F(x) = 1 - e^{-x}, -\infty < x < +\infty$ B. $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x, -\infty < x < +\infty$
- C. $F(x) = \begin{cases} 1 + e^{-x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$ D. $F(x) = \begin{cases} \frac{1}{2}(1 + \cos x), & |x| < \frac{\pi}{2}, \\ 0, & |x| \geq \frac{\pi}{2} \end{cases}$

二、填空题

9. 设电路由元件 A 与 B 并联而成, 且 A 与 B 相互独立, 两元件正常工作的概率都为 p , 已知电路正常工作的概率为 $\frac{15}{16}$, 求 $p =$ _____.

10. 设随机变量 X 服从均匀分布 $U[0, 10]$, 求方程 $x^2 + Xx + 1 = 0$ 有实根的概率为 _____.

11. 设随机变量 X 的期望 $E(X)$ 为一非负值, 且 $E\left(\frac{X^2}{2} - 1\right) = 2$, $D\left(\frac{X}{2} - 1\right) = \frac{1}{2}$, 则 $E(X) =$ _____.

12. 设总体 $X \sim B(1, p)$, p 为未知参数. 设 $X_1, X_2, \dots, X_n$ 是来自该总体的样本, $\bar{X}$ 是它们的样本均值, 则 $P\left\{\bar{X} = \frac{k}{n}\right\} =$ _____.

13. 设随机变量 X 的分布函数为

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

其中 A 为待定常数, 则有 $P\{1 < X < 3\} =$ _____.

14. 已知随机变量 X 服从参数为 λ 的泊松分布, 且 $P\{X=0\} = \frac{1}{2}$, 则 $P\{X < 2\} =$ _____.

三、解答题

15. 一工人看 3 台机床, 在 1 h 内机床不需要照顾的概率分别为 0.9, 0.8, 0.7, 且相互独立. 设 X 为 1 h 内需照顾的机床台数, 求 X 的分布律.

16. 设甲箱中有 a 个白球, b 个红球, 乙箱中有 c 个白球, d 个红球. 从甲箱中任取一球放入乙箱中, 然后再从乙箱中任取一球. 求从乙箱中取到的球为白球的概率.

17. 设 X 为一离散型随机变量, 其分布律为

X	-1	0	1
P	$\frac{1}{2}$	$1-2q$	q^2

求 (1) q 值;

(2) X 的分布函数;

(3) $Y = X^2 + 1$ 的分布律.

18. 设连续型随机变量 X 的密度函数为

$$f(x) = \begin{cases} Ax^2 e^{-kx}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (k \text{ 为已知常数}).$$

求:(1)常数 A ;

$$(2) P\left\{0 \leq X \leq \frac{1}{k}\right\}.$$

19. 设随机变量 X 的密度函数为

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

求随机变量 $Y = 2X^3$ 的概率密度函数.

20. 设随机变量 X 的概率密度函数为

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

求 $E(X)$, $D(X)$.

21. 设总体 X 的概率密度为

$$f(x; \theta) = \begin{cases} \frac{1}{\sqrt{\theta}} e^{-\frac{x-\beta}{\sqrt{\theta}}}, & x \geq \beta, \\ 0, & x < \beta, \end{cases}$$

其中参数 $\theta > 0$ 已知, 参数 β 未知, $X_1, X_2, \dots, X_n$ 是来自 X 的一个样本, 求未知参数 β 的矩估计.

22. 岩石密度的测量结果 $X \sim N(\mu, \sigma^2)$. 现抽取 12 个样品, 测得 $\sum_{i=1}^{12} x_i^2 = 89.92$. 若已知 $\mu = 2.7$, 求方差 σ^2 的置信区间 ($\alpha = 0.1$).

23. 某种导线的电阻(单位: Ω) X 服从正态分布 $N(\mu, 0.005^2)$. 现从新生产的一批导线中抽取 9 根, 测其电阻 $s = 0.08 \Omega$. 对于 $\alpha = 0.05$, 能否认为这批导线电阻的方差仍为 0.005^2 ?

24. 设事件 A 与 B 独立, 证明 $\bar{A}$ 与 $\bar{B}$ 独立.

期末测试(一)参考答案

一、选择题

1.D 2.C 3.C 4.A 5.B 6.C 7.B 8.B

二、解答题

$$9. \frac{3}{4} \quad 10. 0.8 \quad 11. 2 \quad 12. C_n^k p^k (1-p)^{n-k} \quad 13. \frac{3}{4} \quad 14. \frac{1}{2}(1 + \ln 2)$$

三、解答题

$$15. \begin{array}{c|cccc} X & 0 & 1 & 2 & 3 \\ \hline P & 0.504 & 0.398 & 0.092 & 0.006 \end{array}.$$

$$16. \frac{a(c+1) + bc}{(a+b)(c+d+1)}.$$

$$17. (1) q = 1 - \sqrt{\frac{1}{2}}.$$

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

$$(3) \begin{array}{c|cc} Y & 1 & 2 \\ \hline P & \sqrt{2}-1 & 2-\sqrt{3} \end{array}.$$

$$18. (1) A = \frac{k^3}{2}.$$

$$(2) P\left\{0 \leq X \leq \frac{1}{k}\right\} = \int_0^{\frac{1}{k}} \frac{k^3}{2} x^2 e^{-kx} dx = 1 - \frac{5}{2e}.$$

$$19. f_Y(y) = F'_Y(y) = \begin{cases} f_X\left(\sqrt[3]{\frac{y}{2}}\right) \frac{1}{3\sqrt[3]{2y^2}}, & y > 0, \\ 0, & y \leq 0, \end{cases}$$

$$\text{即 } f_Y(y) = \begin{cases} \frac{2\sqrt[3]{2}}{3\pi(y\sqrt[3]{y} + \sqrt[3]{4y^2})}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

$$20. E(X) = \int_0^1 x \cdot x dx + \int_1^2 x(2-x) dx = 1,$$

$$D(X) = E(X^2) - E^2(X) = \frac{7}{6} - 1 = \frac{1}{6}.$$

$$21. \beta \text{ 的矩估计为 } \hat{\beta} = \bar{X} - \sqrt{\theta}.$$

$$22. \text{置信区间为 } (0.12, 0.47).$$

$$23. H_0: \sigma^2 = \sigma_0^2 = 0.005^2, H_1: \sigma^2 \neq 0.005^2.$$

拒绝 H_0 , 即不能认为这批导线电阻的方差是 0.005^2 .

24. 由 $P(AB) = P(A) \cdot P(B)$ 得

$$\begin{aligned} P(\bar{A})P(\bar{B}) &= [1 - P(A)][1 - P(B)] \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= 1 - P(A) - P(B) + P(AB) \\ &= 1 - P(A \cup B) \\ &= P(\overline{A \cup B}) = P(\bar{A}\bar{B}). \end{aligned}$$

因此 $\bar{A}$ 与 $\bar{B}$ 也独立.

期末测试(二)

一、选择题

1. 设 A 表示事件“甲种产品滞销, 乙种产品滞销”, 则其逆事件 $\bar{A}$ 为().

A. “甲种产品滞销, 乙种产品畅销”

B. “甲、乙两种产品均畅销”

C. “甲种产品滞销”

D. “甲种产品或乙种产品畅销”

2. 设随机变量 X 与 Y 均服从正态分布, $X \sim N(\mu, A^2)$, $Y \sim N(\mu, S^2)$. 记 $p_1 = P\{X \leq \mu - 4\}$, $p_2 = P\{Y \geq \mu + 5\}$, 则().

- A. 对任何实数 μ , 都有 $p_1 = p_2$ B. 对任何实数 μ , 都有 $p_1 < p_2$
C. 只对 μ 的个别值, 才有 $p_1 = p_2$ D. 对任何实数 μ , 都有 $p_1 > p_2$

3. 对任意两个随机变量 X 和 Y , 若 $E(XY) = EX \cdot EY$, 则().

- A. $D(XY) = DX \cdot DY$ B. $D(X + Y) = DX + DY$
C. X 与 Y 相互独立 D. X 与 Y 不独立

4. 设 $\{X_n\}$ 是独立同分布的随机变量序列, $EX_n = \mu$, 则对于任意给定的 $\varepsilon > 0$, 有().

- A. $\lim_{n \rightarrow \infty} P\left\{\frac{1}{n} \left| \sum_{i=1}^n X_i - n\mu \right| \geq \varepsilon\right\} = 0$ B. $\lim_{n \rightarrow \infty} P\left\{\frac{1}{n} \left| \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon\right\} = 0$
C. $\lim_{n \rightarrow \infty} P\left\{\frac{1}{n} \left| \sum_{i=1}^n X_i - n\mu \right| \leq \varepsilon\right\} = 0$ D. $\lim_{n \rightarrow \infty} P\left\{\frac{1}{n} \left| \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon\right\} = 1$

5. 设 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 都是参数 θ 的无偏估计量, $\hat{\theta} = a\hat{\theta}_1 + b\hat{\theta}_2$ 也是参数 θ 的无偏估计量, 则 a, b 分别为().

- A. $a = \frac{1}{2}, b = -\frac{1}{2}$ B. $a = \frac{2}{3}, b = \frac{1}{6}$ C. $a = \frac{3}{2}, b = -\frac{1}{2}$ D. $a = \frac{1}{3}, b = \frac{1}{2}$

二、填空题

6. 设 A, B 是随机事件, $P(A) = 0.7$, $P(A - B) = 0.3$, 则 $P(\overline{AB}) =$ _____.

7. 设随机变量 X 服从参数 $\lambda = 1$ 的指数分布, 则 $E(X + e^{-2X}) =$ _____.

8. 设总体 X 的概率密度函数为 $p(x)$, $(X_1, X_2, \dots, X_n)$ 是从中抽取的一个样本, 则 $(X_1, \dots, X_n)$ 的联合密度函数为 _____.

9. 设总体 X 与 Y 相互独立, 且都服从正态分布 $N(0, 1)$, $(X_1, \dots, X_9)$ 是从总体 X 中抽取的一个样本, $(Y_1, \dots, Y_9)$ 是从总体 Y 中抽取的一个样本, 则统计量 $U = \frac{X_1 + X_2 + \dots + X_9}{\sqrt{Y_1^2 + Y_2^2 + \dots + Y_9^2}}$ 服从 _____ 分布, 参数为 _____.

10. 连续型随机变量 X 的概率密度为 $f(x) = Ae^{-|x|} (-\infty < x < +\infty)$, 则 $A =$ _____.

三、解答题

11. 在调查某城市居民的消费水平时, 发现该城市家庭中有 15% 已购买空调, 12% 已购买电脑, 20% 已购买 VCD 机, 6% 的家庭已购买空调和电脑, 10% 的家庭已购买空调和 VCD 机, 5% 的家庭已购买电脑和 VCD 机, 三种电器都已购买的为 2%. 试求下列事件的概率: (1) 只购买空调的; (2) 只购买两种电器的.

12. 甲、乙、丙 3 台机床加工同一种零件, 零件由各台机床加工的百分比依次是 50%、30%、20%. 各机床加工的优质品率依次是 80%、85%、90%, 将加工的零件混在一起, 从中任取 1 个, 求取得优质品的概率.

13. 两台相同型号的自动记录仪, 每台无故障工作的时间分别为 X 和 Y , 假设 X 与 Y 相互独立, 都服从参数为 $\lambda = 5$ 的指数分布. 现首先开动其中一台, 当其损坏停用时另一台自动开动, 直至第二台记录仪损坏为止. 令 T 表示从开始到第二台记录仪损坏时记录仪的总共工作时间, 试求随机变量 T 的概率密度函数.

14. 设二维随机变量 (X, Y) 的联合密度函数为

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

试求:(1)常数 A ;(2)随机变量 X 的边缘密度函数 $f_X(x)$.

15. 罐中有 5 个红球, 3 个白球, 有放回地每次取一球, 直到取得红球为止. 用 X 表示抽取次数, 求 X 的分布律, 并计算 $P\{1 \leq X \leq 3\}$.

16. 假设总体 X 服从正态分布 $N(1, 0.2^2)$, 样本 $X_1, \dots, X_n$ 来自总体 X , 要使样本均值 $\bar{X}$ 满足概率不等式

$$P\{0.9 \leq \bar{X} \leq 1.1\} \geq 0.95,$$

求样本容量 n 最小应取多大?

17. 设总体 X 的密度函数为 $p(x; \theta, \lambda) = \frac{1}{2\lambda} e^{-\frac{|x-\theta|}{\lambda}}$, 其中 $\lambda > 0$, 设 $x_1, x_2, \dots, x_n$ 是来自这一总体的样本, 求 θ, λ 的矩估计.

18. 设随机变量 $X \sim N(\mu, 2.8^2)$ 现有 X 的 10 个观察值 $x_1, \dots, x_{10}$, 已知 $\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 1.500$, 求:(1) μ 的置信度为 0.95 的置信区间;(2)要使 0.95 的置信区间长度小于 1, 观察值个数 n 最少应取多少?

19. 设某市青少年犯罪的年龄构成服从正态分布, 今随机抽取 9 名罪犯, 其年龄如下:

22, 17, 19, 25, 25, 18, 16, 23, 24

试以 95% 的概率判断犯罪青少年的平均年龄是否为 18 岁.

20. 设 A, B 是两个事件, 又设 $P(A) = p_1 > 0, P(B) = p_2 > 0$ 且 $p_1 + p_2 > 1$. 证明:

$$P(B|A) \geq 1 - \frac{1-p_2}{p_1}.$$

期末测试(二)参考答案

一、选择题

1. D 2. A 3. B 4. A 5. C

二、填空题

6. $\frac{3}{5}$ 7. $\frac{4}{3}$ 8. $\prod_{i=1}^n p(x_i)$ 9. $t/9$ 10. $\frac{1}{2}$

三、解答题

11. (1) $P\{\text{某家庭只购买空调}\} = 0.01$.

(2) $P\{\text{某家庭只购买两种电器}\} = 0.15$.

12. 设 D 表示取出的产品是优质品.

$$P(D) = 0.835.$$

13. 随机变量 T 的密度函数为 $p_T(t) = \begin{cases} 25te^{-5t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$

14. (1) $A = 24$.

(2) 随机变量 X 的边缘密度函数为 $f_X(x) = \begin{cases} 12x^2(1-x), & 0 \leq x \leq 1, \\ 0, & \text{其他.} \end{cases}$

$$15. \begin{array}{c|cccccc} X & 1 & 2 & 3 & \dots & k & \dots \\ \hline p & \frac{5}{8} & \frac{15}{64} & \frac{45}{512} & \dots & \left(\frac{3}{8}\right)^{k-1} \frac{5}{8} & \dots \end{array}$$

故 $P\{1 < X \leq 3\} = P\{X=2\} + P\{X=3\} = \frac{15}{64} + \frac{45}{512} \approx 0.322$.

16. 样本容量 n 最少应取为 16.

$$17. \begin{cases} \hat{\theta} = \bar{x}, \\ \hat{\lambda} = \frac{\hat{\sigma}}{\sqrt{2}} = \sqrt{\frac{1}{2n} \sum_{i=1}^n (x_i - \bar{x})^2}. \end{cases}$$

18. (1) 置信区间为 (1 498.265, 1 501.735)

(2) 样本容量 n 至少应取 121.

19. 检验假设 $H_0: \mu = 18, H_1: \mu \neq 18$.

拒绝 H_0 .

$$20. P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(A) + P(B) - P(A+B)}{P(A)},$$

由于 $P(A+B) \leq 1$, 所以

$$P(A+B) \geq \frac{p_1 + p_2 - 1}{p_1} = 1 - \frac{1 - p_2}{p_1}.$$

期末测试(三)

一、选择题

1. 设 A, B, C 为随机事件, 若 $P(C) > 0$ 且

$$P(A \cup B | C) = P(A | C) + P(B | C)$$

则下列不等式成立的是().

$$A. P(C(A \cup B)) = P(AC) + P(BC) \quad B. P(A \cup B) = P(A | C) + P(B | C)$$

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

$$D. P(A \cup B | \bar{C}) = P(A | \bar{C}) + P(B | \bar{C})$$

2. 设甲、乙二人独立地向同一目标各射击 1 次, 其命中率分别为 0.6 和 0.5, 现两人同时射击各一次, 则目标被击中的概率是().

$$A. 0.1$$

$$B. \frac{3}{11}$$

$$C. 0.8$$

$$D. 0.6$$

3. 设某连续型随机变量的密度函数 $p(x)$ 为偶函数, $F(x)$ 为其分布函数, 则对任意实数 a , 有().

$$A. F(-a) = 1 - \int_0^a p(x) dx$$

$$B. F(-a) = 2F(a) - 1$$

$$C. F(-a) = \frac{1}{2} - \int_0^a p(x) dx$$

$$D. F(-a) = F(a)$$

4. 设 $E(X) = 1, E(Y) = 2, D(X) = 1, D(Y) = 4, \rho_{XY} = 0.6$, 则 $E(2X - Y + 1)^2 =$ ().

$$A. 2.1$$

$$B. 3.2$$

$$C. 4.2$$

$$D. 4.8$$

5. 设随机变量 X 的方差 $D(X) < +\infty$, a 为正常数, 则 $P\left(\frac{|X - E(X)|}{a} \geq 1\right) \leq$ ().

- A. $D(X)$ B. $1 - D(X)$ C. $1 - \frac{D(X)}{a^2}$ D. $\frac{D(X)}{a^2}$

6. 设 (X, Y) 的联合密度为 $p(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & \text{其他} \end{cases}$, 则 X 与 Y 为 () 的随机变量.

- A. 不独立同分布 B. 独立不同分布 C. 独立同分布 D. 不独立也不同分布

7. 设总体 X 和 Y 相互独立, 分别服从 $N\left(0, \frac{1}{n}\right)$ 和 $N\left(0, \frac{1}{m}\right)$, n 和 m 是自然数, 且 $n \neq m$. $X_1, X_2, \dots, X_k$ 和 $Y_1, Y_2, \dots, Y_l$ 分别是 X 和 Y 的样本, 要使 $\frac{X_1 + X_2 + \dots + X_k}{\sqrt{Y_1^2 + Y_2^2 + \dots + Y_l^2}}$ 服从 t 分布, k, l 应满足条件 ().

- A. $k = l$ B. $\frac{k}{l} = \frac{n}{m}$ C. $kl = nm$ D. $\frac{k}{l} = \frac{m}{n}$

8. 样本 $X_1, X_2, \dots, X_n (n \geq 3)$ 取自总体 X , 则下列估计量中, 不是总体期望 μ 的无偏估计量有 ().

- A. $\bar{X}$ B. $X_1 + X_2 + \dots + X_n$ C. $0.1(6X_1 + 4X_n)$ D. $X_1 + X_2 - X_3$

9. 总体均值 μ 的区间估计中, 正确的说法是 ().

- A. 置信度 $1 - \alpha$ 一定时, 样本容量增加, 则置信区间的长度变长
B. 置信度 $1 - \alpha$ 一定时, 样本容量增加, 则置信区间的长度变短
C. 置信度 $1 - \alpha$ 变小, 则置信区间的长度变短
D. 置信度 $1 - \alpha$ 变大, 则置信区间的长度变短

10. 在假设检验中, 显著性水平 α 是指 ().

- A. $P(\text{接受 } H_0 | H_0 \text{ 为假}) = \alpha$ B. $P(\text{接受 } H_1 | H_1 \text{ 为假}) = \alpha$
C. $P(\text{拒绝 } H_0 | H_0 \text{ 为真}) = \alpha$ D. $P(\text{拒绝 } H_1 | H_1 \text{ 为真}) = \alpha$

二、填空题

11. 已知 $P(A \cup B) = 0.8$, $P(B) = 0.4$, 则 $P(A | \bar{B}) =$ _____.

12. 某种商品一天的销售量(件)服从参数为 λ 的泊松分布, 随机选取 4 天, 其中恰有一天的销售量为 5 件的概率是 _____.

13. 设相互独立的随机变量 X 与 Y 具有相同的分布, 且 X 的分布律为 $\begin{matrix} X & -1 & 1 & 2 \\ P & 0.2 & 0.2 & 0.6 \end{matrix}$, 则随机变量 $Z = \max\{X^2, Y^2\}$ 的分布律为 _____.

14. 用切比雪夫不等式估计: 投掷一枚均匀的硬币 100 次, 正面向上出现的频率在 0.4 到 0.6 之间的概率不小于 _____.

15. 设 $X_1, X_2, \dots, X_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的样本, 已知 $\sigma^2 = \sigma_0^2$, 检验假设 $H_0: \mu = \mu_0$, 则当 H_0 为真时, 常选用的统计量是 _____, 它服从的分布为 _____.

三、解答题

16. 设有两台机床加工同样的零件, 第一台机床出废品的概率是 0.03, 第二台机床出废品

的概率是 0.02 加工出来的零件混放在一起,并且已知第一台机床的加工的零件比第二台机床多一倍.

(1)求任取一个零件是合格品的概率;

(2)若任取的一个零件经检查后发现有废品,则它是第二台机床加工的概率.

17. 设随机变量 (X, Y) 的联合分布律:

$Y \backslash X$	0	1	2
0	0.125	0.25	0
2	0.125	0.25	0.25

求 (1) $Z = XY$ 的分布律;

(2) 条件分布律 $P(X = i | Y = 2)$.

18. 设二维随机变量 (X, Y) 的密度函数为

$$f(x, y) = \begin{cases} ce^{-(3x+4y)}, & x, y > 0, \\ 0, & \text{其他.} \end{cases}$$

(1) 确定常数 c ;

(2) 讨论 X, Y 的独立性;

(3) 计算 $P(0 < X \leq 1, 0 < Y \leq 2)$.

19. 一通信系统拥有 50 台相互独立起作用的交换机. 在系统运行期间, 每台交换机能清晰接受信号的概率为 0.90, 系统正常工作时, 要求能清晰接受信号的交换机至少 45 台. 求该通信系统能正常工作的概率.

20. 设随机变量 X 服从 $(0, \theta)$ 上的均匀分布, $X_1, X_2, \dots, X_n$ 是从总体 X 中抽取的简单随机样本, 求 θ 的矩估计量 $\hat{\theta}_1$ 和极大似然估计量 $\hat{\theta}_2$.

21. 从理论分析得出结论: 压缩机的冷却用水, 其温度服从 $X \sim N(\mu, \sigma^2)$, 升高的平均值不多于 5°C . 现测量了 5 台压缩机的冷却用水的升高温度分别是 6.4 4.3 5.7 4.9 5.4. 问在 $\alpha = 0.05$ 时, 这组数据与理论分析得出的结论是否一致?

22. 设连续随机变量 X 的密度函数 $f(x)$ 在 $x < 0$ 时恒为零, 且 $E(X)$ 存在. 试证: 对任意正常数 a 有 $P(X > a) \leq \frac{E(X)}{a}$.

23. 证明在一次试验中, 事件 A 发生的次数 X 的方差 $D(X) \leq \frac{1}{4}$.

期末测试(三)参考答案

一、选择题

1. A 2. C 3. C 4. C 5. D 6. A 7. B 8. B 9. B 10. C

二、填空题

11. $\frac{2}{3}$ 12. $4\left(\frac{\lambda^5}{5!}e^{-\lambda}\right)\left(1 - \frac{\lambda^5}{5!}e^{-\lambda}\right)^3$ 13. $\frac{Z}{P} \left| \begin{matrix} 1 & 4 \\ 0.16 & 0.84 \end{matrix} \right|$ 14. $\frac{3}{4}$

15. $\frac{\bar{X} - \mu_0}{\sigma_0} \sqrt{n} \sim N(0, 1)$

三、解答题

16. 设 A_i 表示零件是第 i 台机床加工的, $i=1, 2$. 事件 B 表示任取的一个零件是合格品, 则

$$(1) P(B) = 0.973.$$

$$(2) P(A_2 | \bar{B}) = 0.25.$$

$$17. (1) \begin{array}{c|ccc} Z & 0 & 2 & 4 \\ \hline P & 0.5 & 0.25 & 0.25 \end{array}$$

$$(2) \begin{array}{c|ccc} X=i | Y=2 & 0 & 1 & 2 \\ \hline P & 0.2 & 0.4 & 0.4 \end{array}.$$

$$18. (1) c = 12.$$

$$(2) f_X(x) = \int_0^{+\infty} 12e^{-(3x+4y)} dy = 3e^{-3x}, \quad x \geq 0,$$

$$f_X(x) = \begin{cases} 3e^{-3x}, & x \geq 0 \\ 0, & x < 0. \end{cases} \quad \text{同理 } f_Y(y) = \begin{cases} 4e^{-4y}, & y \geq 0 \\ 0, & y < 0. \end{cases}$$

等式 $f(x, y) = f_X(x)f_Y(y)$ 对任何 (x, y) 成立. 故 X 与 Y 相互独立.

$$(3) P(0 < X \leq 1, 0 < Y \leq 2) = \int_0^1 3e^{-3x} dx \int_0^2 4e^{-4y} dy = (1 - e^{-3})(1 - e^{-4}).$$

19. 设 X 表示系统运行期间能清晰接受信号的交换机台数, 则 $X \sim B(50, 0.9)$.

$$P(\text{通信系统能正常工作}) = P(45 \leq X \leq 50)$$

$$= P\left(\frac{45 - 50 \times 0.9}{\sqrt{50 \times 0.9 \times 0.1}} \leq \frac{X - 50 \times 0.9}{\sqrt{50 \times 0.9 \times 0.1}} \leq \frac{50 - 50 \times 0.9}{\sqrt{50 \times 0.9 \times 0.1}}\right)$$

$$= \Phi(2.36) - \Phi(0) = 0.4909.$$

$$20. \hat{\theta}_1 = 2\bar{X}.$$

$$\hat{\theta}_2 = X_{(n)} = \max\{X_1, X_2, \dots, X_n\}.$$

$$21. H_0: \mu = 5; H_1: \mu > 5.$$

接受 H_0 , 即这组数据与理论分析得出的结论一致.

$$22. P(X > a) = \int_a^{+\infty} f(x) dx \leq \frac{1}{a} \int_a^{+\infty} xf(x) dx \leq \frac{1}{a} \int_0^{+\infty} xf(x) dx$$

$$= \frac{1}{a} \int_{-\infty}^{+\infty} xf(x) dx = \frac{1}{a} E(X).$$

23. 设 $P(A) = p$, 则 X 的分布律为

$$\begin{array}{c|cc} X & 0 & 1 \\ \hline P & 1-p & p \end{array}.$$

$$E(X) = p, E(X^2) = p, \text{ 从而 } D(X) = E(X^2) - [E(X)]^2 = p - p^2 = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leq \frac{1}{4}.$$

期末测试(四)

一、选择题

1. 若二事件 A 和 B 同时出现的概率 $P(AB) = 0$, 则().

- A. A 和 B 不相容(相斥) B. AB 是不可能事件
C. AB 未必是不可能事件 D. $P(A)=0$ 或 $P(B)=0$

2. 若 A, B 为两个随机事件, 且 $B \subset A$, 则下列式子正确的是().

- A. $P(A \cup B) = P(A)$ B. $P(AB) = P(A)$
C. $P(B|A) = P(B)$ D. $P(B - A) = P(B) - P(A)$

3. 设随机变量 X 服从正态分布 $N(\mu, \sigma^2)$, 则随 σ 的增大, 概率 $P\{|X - \mu| < \sigma\}$ ().

- A. 单调增大 B. 单调减少 C. 相关系数不为零 D. 相关系数为零

4. 设 $F_1(x)$ 与 $F_2(x)$ 分别为随机变量 X_1 与 X_2 的分布函数, 为使 $F(x) = aF_1(x) - bF_2(x)$ 是某一随机变量的分布函数, 在下列给定的各组数值中应取().

- A. $a = \frac{3}{5}, b = -\frac{2}{5}$ B. $a = \frac{2}{3}, b = \frac{2}{3}$ C. $a = -\frac{1}{2}, b = \frac{3}{2}$ D. $a = \frac{1}{2}, b = -\frac{3}{2}$

5. 假设随机变量 X 服从指数分布, 则随机变量 $Y = \min(X, 2)$ 的分布函数().

- A. 是连续函数 B. 至少有两个间断点 C. 是阶梯函数 D. 恰好有一个间断点

6. 设 X 是一随机变量, $E(X) = \mu, D(X) = \sigma^2$ ($\mu, \sigma > 0$ 常数), 则对任意常数 C , 必有().

- A. $E(X - C)^2 = EX^2 - C^2$ B. $E(X - C) = E(X - \mu)^2$
C. $E(X - C)^2 < E(X - \mu^2)$ D. $E(X - C)^2 \geq E(X - \mu)^2$

7. 设随机变量 X 和 Y 相互独立, 其概率分布为

X	-1	1
P	$\frac{1}{2}$	$\frac{1}{2}$

,

Y	-1	1
P	$\frac{1}{2}$	$\frac{1}{2}$

则下列式子正确的是().

- A. $X = Y$ B. $P\{X = Y\} = 0$ C. $P\{X = Y\} = \frac{1}{2}$ D. $P\{X = Y\} = 1$

8. 设随机变量 $X_1, X_2, \dots, X_9$ 相互独立且同分布, 而且有 $EX_i = 1, DX_i = 1$ ($i = 1, 2, \dots, 9$), 令 $X = \sum_{i=1}^9 X_i$, 则对任意给定的 $\epsilon > 0$, 由切比雪夫不等式直接可得().

- A. $P\{|X - 1| < \epsilon\} \geq 1 - \frac{1}{\epsilon^2}$ B. $P\{|X - 9| < \epsilon\} \geq 1 - \frac{1}{\epsilon^2}$
C. $P\{|X - 9| < \epsilon\} \geq 1 - \frac{9}{\epsilon^2}$ D. $P\{|\frac{1}{9}X - 1| < \epsilon\} \geq 1 - \frac{1}{\epsilon^2}$

9. 设 $X \sim N(0, 16), Y \sim N(0, 9)$, X, Y 相互独立, $X_1, X_2, \dots, X_9$ 与 $Y_1, Y_2, \dots, Y_{16}$ 分别为 X 与 Y 的一个简单随机样本, 则 $\frac{X_1^2 + X_2^2 + \dots + X_9^2}{Y_1^2 + Y_2^2 + \dots + Y_{16}^2}$ 服从的分布为().

- A. $F(9, 16)$ B. $F(16, 9)$ C. $F(9, 9)$ D. $F(16, 16)$

10. 设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(0, \sigma^2)$ 的简单随机样本, 则 σ^2 的无偏估计量是().

- A. $\frac{1}{n} \sum_{i=1}^n X_i^2$ B. $\frac{1}{n-1} \sum_{i=1}^n X_i^2$ C. $\frac{1}{n^2} \sum_{i=1}^n X_i^2$ D. $\frac{1}{n+1} \sum_{i=1}^n X_i^2$

二、填空题

11. 设两个相互独立的事件 A 和 B 都不发生的概率为 $\frac{1}{9}$, A 发生 B 不发生的概率与 B 发生 A 不发生的概率相等, 则 $P(A) = \underline{\hspace{2cm}}$.

12. 设随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{3}, & x \in [0, 1], \\ \frac{2}{9}, & x \in [3, 6], \\ 0, & \text{其他.} \end{cases}$$

若 k 使得 $P(X \geq k) = \frac{2}{3}$, 则 k 的取值范围是 $\underline{\hspace{2cm}}$.

13. 设二维随机变量 (X, Y) 的概率密度为

$$f(x, y) = \begin{cases} 6x, & 0 \leq x \leq y \leq 1, \\ 0, & \text{其他,} \end{cases}$$

则 $P\{X + Y \leq 1\} = \underline{\hspace{2cm}}$.

14. 已知一批零件的长度 X (单位: cm) 服从正态分布 $N(\mu, 1)$, 从中随机地抽取 16 个零件, 得到长度的平均值为 40 (cm), 则 μ 的置信度为 0.95 的置信区间是 $\underline{\hspace{2cm}}$.

(注: 标准正态分布函数值 $\Phi(1.96) = 0.975$, $\Phi(1.645) = 0.95$)

15. 设随机变量 X 和 Y 的联合概率分布为

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

则 X^2 和 Y^2 的协方差 $\text{cov}(X^2, Y^2) = \underline{\hspace{2cm}}$.

三、解答题

16. 假设有两箱同种零件: 第一箱内装 50 件, 其中 10 件一等品, 第二箱内装 30 件, 其中 18 件一等品. 现从两箱中随意挑出一箱, 然后从该箱中先后随机取出两个零件 (取出的零件均不放回). 试求:

(1) 先取出的零件是一等品的概率 p ;

(2) 在先取出的零件是一等品的条件下, 第二次取出的零件仍然是一等品的条件概率 q .

17. 设二维随机变量 (X, Y) 的概率密度为

$$f(x, y) = \begin{cases} 2e^{-(x+2y)}, & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases}$$

求随机变量 $Z = X + 2Y$ 的分布函数.

18. 设总体 X 的概率密度为

$$f(x) = \begin{cases} 2e^{-2(x-\theta)}, & x > \theta, \\ 0, & x \leq \theta, \end{cases}$$

其中 $\theta > 0$ 是未知参数. 从总体 X 中抽取简单随机样本 $X_1, X_2, \dots, X_n$, 记 $\hat{\theta} = \min(X_1, X_2, \dots, X_n)$.

- (1)求总体 X 的分布函数 $F(x)$;
- (2)求统计量 $\hat{\theta}$ 的分布函数 $F_{\hat{\theta}}(x)$;
- (3)如果用 $\hat{\theta}$ 作为 θ 的估计量,讨论它是否具有无偏性.

19. 一商店经销某种商品,每周进货的数量 X 与顾客对该种商品的需求量 Y 是相互独立的随机变量,且都服从区间 $[10, 20]$ 上的均匀分布.商店每售出一单位商品可获得利润 1 000 元,若需求量超过了进货量,商店可以从其他商店调剂供应,这时每单位商品获利润 500 元,试计算此商店经销该种商品每周所得利润的期望值.

20. 设随机变量 X 的分布函数为

$$F(x; \alpha, \beta) = \begin{cases} 1 - \left(\frac{\alpha}{x}\right)^\beta, & x > \alpha, \\ 0, & x \leq \alpha, \end{cases}$$

其中参数 $\alpha > 0, \beta > 1$. 设 $X_1, X_2, \dots, X_n$ 为来自总体 X 的简单随机样本.

- (1)当 $\alpha = 1$ 时,求未知参数 β 的矩估计量;
- (2)当 $\alpha = 1$ 时,求未知参数 β 的极大似然估计量;
- (3)当 $\beta = 2$ 时,求未知参数 α 的极大似然估计量.

21. 假设 0.50, 1.25, 0.80, 2.00 是来自总体 X 的简单随机样本值,已知 $Y = \ln X$ 服从正态分布 $N(\mu, 1)$.

- (1)求 X 的数字期望 $E(X)$ (记 $E(X)$ 为 b);
- (2)求 μ 的置信度为 0.95 的置信区间;
- (3)利用上述结果求 b 的置信度为 0.95 的置信区间.

22. 设某次考试的考生成绩服从正态分布,从中随机地抽取 36 位考生的成绩,平均成绩为 66.5 分,标准差为 15 分.问在显著性水平 0.05 下,是否可以认为这次考试全体考生的平均成绩为 70 分?并给出检验过程.

期末测试(四)参考答案

一、选择题

1.C 2.A 3.C 4.A 5.D 6.D 7.C 8.C 9.A 10.B

二、填空题

11. $\frac{2}{3}$ 12. $[1, 3]$ 13. $\frac{1}{4}$ 14. (39.51, 40.49) 15. -0.02

三、解答题

$$16. (1) p = 0.4 \quad (2) q = 0.485 \quad 6. \quad 17. F_Z(z) = \begin{cases} 1 - e^{-z} - ze^{-z}, & z > 0, \\ 0, & z \leq 0. \end{cases}$$

$$18. (1) F(x) = \begin{cases} 1 - e^{-2(x-\theta)}, & x > \theta, \\ 0, & x \leq \theta. \end{cases} \quad (2) F_{\hat{\theta}}(x) = \begin{cases} 1 - e^{-2n(x-\theta)}, & x > \theta, \\ 0, & x \leq \theta. \end{cases}$$

(3) $E(\hat{\theta}) = \theta + \frac{1}{2n} \neq \theta$, $\hat{\theta}$ 作为 θ 的估计量不具有无偏性.

$$19. 14 \quad 166.67(\text{元}). \quad 20. (1) \hat{\beta} = \frac{\bar{X}}{\bar{X} - 1}. \quad (2) \hat{\beta} = \frac{n}{\sum_{i=1}^n \ln X_i}. \quad (3) \hat{\alpha} = \min\{X_1, X_2, \dots, X_n\}.$$

$$21. (1) E(X) = e^{\mu + \frac{1}{2}}. \quad (2) (-0.98, 0.98). \quad (3) (e^{-0.48}, e^{1.48}). \quad 22. H_0: \mu = 70 \text{ 接受 } H_0.$$