高职、高专教学辅导用书

概率论与数理统计
学习指导与习题解析

贺伟奇　涂德胜　裘亚峥　编

华中科技大学出版社

图书在版编目(CIP)数据

概率论与数理统计学习指导与习题解析/ 贺伟奇　涂德胜　裘亚峥　编
武汉:华中科技大学出版社
ISBN 7-5609-3107-3
Ⅰ.概…
Ⅱ.① 贺…　② 涂…　③ 裘…
Ⅲ.
Ⅳ.O21

概率论与数理统计学习指导与习题解析

贺伟奇
涂德胜　编
裘亚峥

责任编辑:叶　兰　　**封面设计:**
责任校对:封春英　　**监　　印:**

出版发行者:华中科技大学出版社　　**武昌喻家山**
邮编:430074　　**电话**:(027)87542624

经销者:新华书店湖北发行所

录排者:华中理工大学出版社照排室排版
印刷者:

开本:　　**印张:**　　**字数:**
版次:　年　月第　版　版次:　年　月第　次印刷　　**印数:**
ISBN 7-5609-　/　·　　**定价:　元**

(本书若有印装质量问题,请向出版社发行科调换)

内 容 提 要

本书是学习概率论与数理统计课程的辅导教材，每章由内容提要、要点范例、习题选解和试题选解四部分组成。其中习题选自全国高职、高专教材《工程数学——概率论与数理统计》一书，试题选自近五年的考试试题。

本书选题突出重点、难点，解题注重思路分析与知识的综合运用，可供各类理工科院校的高职、高专学生使用，也可供有关教师参考。

前　　言

“概率论与数理统计”是理工科学生的一门必修课。由于该学科与社会、经济以及自然现象联系密切,在工程技术和科学实验中有着广泛的应用,因此学好这一课程是十分重要的。

在多年的教学中,我们深深感到,学生学习本课程普遍觉得吃力,表现在对某些概念难以准确掌握;做习题时感觉茫然,甚至解完题后也不知正确与否。

针对这一状况,我们编写了这本辅导教材,想帮助读者(特别是自学者)从以下几个方面来学好本课程。

1.掌握“三基”:在内容提要中,我们通过知识脉络和知识要点两种形式对每章的三基——基本概念、基本理论和基本方法进行了归纳总结,并揭示其内在联系。

2.突破难点:我们对每章都归纳出几个主要的题型,精心选配了典型例题进行解题分析、详解和评注,以帮助学生化解疑难。

3.综合运用:我们通过对习题、试题的选解,从思路分析和解题方法两方面来培养学生综合运用所学知识的能力。

应当指出,学习任何一门课程,学习者独立的思考和自身的感悟是非常重要的,任何好的辅导书都不能替代这些。

在本书出版时,要特别感谢周泰文老师的精心策划和肖果能教授的认真审稿。由于编者水平有限,不妥之处,望读者指正。

编　者

2003 年 10 月

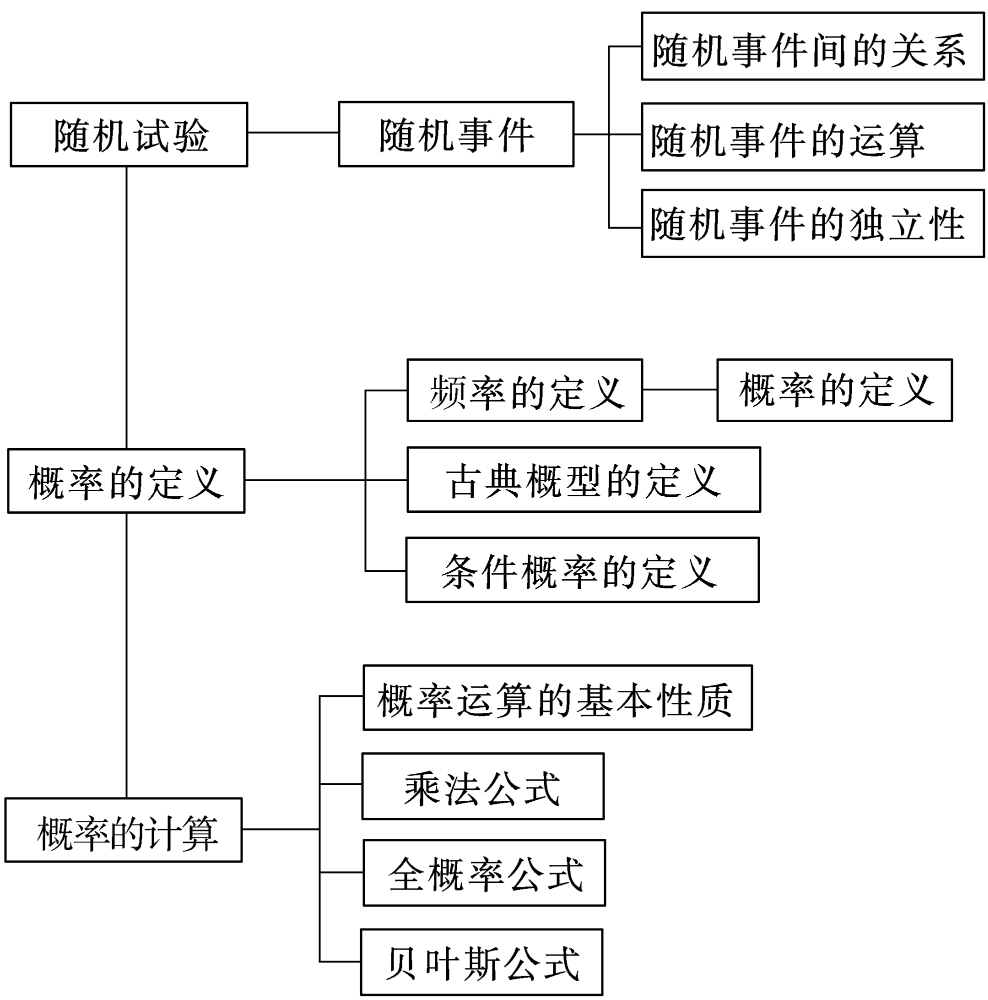
于中南大学铁道学院

目　　录

第一章　随机事件与概率

内 容 提 要

一、知识脉络



二、知识要点

(一) 随机试验

满足下列条件的试验称为随机试验.

(1) 试验可以在相同条件下重复进行;

(2) 每次试验结果不止一个,并能事先明确试验的所有可能结果;

(3) 试验前不知道会出现什么结果.

(二) 随机事件

在随机试验中,可能发生也可能不发生的事件称为随机事件(简称为事件).一般用大写字母 A, B, C 等表示.必然会发生的事件叫做必然事件,用记号 Ω 表示.必然不发生的事件叫做不可能事件,用记号　表示.

(三) 事件间的关系

(1) 若事件 A 发生必然导致事件 B 发生,则称事件 B 包含事件 A,记为 B　A.

(2) 若事件 A 和 B 满足 A　B 和 B　A,称事件 A, B 相等,记为 $A = B$.

(3) 事件 A 与 B 中至少有一个发生的事件称为事件 A 与 B 的和(或并),记为 $A \cup B$(或 $A + B$).

(4) 事件 A 与 B 同时发生的事件称为事件 A 与 B 的积(或交),记为 $A \cap B$(或 AB).

(5) 事件 A 发生而事件 B 不发生的事件称为事件 A 与 B 的差,记为 $A - B$(或 $A\overline{B}$).

(6) 若事件 A 与 B 不能同时发生,即 $AB =$　,则称事件 A 与 B 互不相容(或称互斥).

(7) 若事件 A 与 B 至少发生一个,又互不相容.即: $A \cup B = \Omega$, $AB =$　,则称 A, B 是互为对立事件(或互逆).这时 $B = \overline{A}$, $A = \overline{B}$.

(四) 事件的运算

(1) 交换律 $A\cup B=B\cup A$, $A\cap B=B\cap A$;

(2) 结合律 $(A\cup B)\cup C=A\cup(B\cup C)$, $(AB)C=A(BC)$;

(3) 分配律 $(A\cup B)C=AC\cup BC$,

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C);$$

(4) 对偶律 $\overline{A\cup B}=\overline{A}\overline{B}$ $\quad \overline{AB}=\overline{A}\cup\overline{B}$.

(五) 古典概型

具有下面两个特点的随机试验的数学模型称为古典概型 .

(1) 随机试验只有有限个基本事件;

(2) 每一个基本事件发生的可能性相同 .

古典概型中随机事件 A 的概率定义为

$$P(A)=\frac{A\text{ 中包含的基本事件数 }M}{\text{基本事件总数 }N}.$$

(六) 概率的性质

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

(2) 可加性:若随机事件 $A_1, A_2, \cdots, A_n$ 互不相容,则

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

(3) $P(\overline{A})=1-P(A)$;

(4) 若 $A\quad B$,则 $P(B-A)=P(B)-P(A)$, $P(A)\leq P(B)$;

(5) 对于任意两个事件 A, B 有

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

设 A, B, C 为任意三个事件,则

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

(七) 条件概率与乘法公式

1 . 条件概率

设 A 与 B 是两个随机事件,且 $P(B)>0$,则规定在事件 B 发

生的条件下事件 A 发生的条件概率为

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

2．乘法公式

若 $P(B)>0$，有

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

若 $P(AB)>0$，则有

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

对任意 n 个事件 $A_1,A_2,\cdots,A_n$，若 $P(A_1A_2\cdots A_{n-1})>0$，则有 $P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)\cdot\cdots\cdot P(A_n|A_1A_2\cdots A_{n-1})$.

（八）全概率公式与贝叶斯公式

1．全概率公式

若事件 $B_1,B_2,\cdots,B_n$ 为完备事件组，且 $P(B_i)>0\ (i=1,2,\cdots,n)$，则对任一事件 A，有

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

2．贝叶斯公式

设 $B_1,B_2,\cdots,B_n$ 为完备事件组，A 为任一事件，且 $P(B_i)>0\ (i=1,2,\cdots,n)$，$P(A)>0$，则

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)}\quad(i=1,2,\cdots,n).$$

（九）事件的独立性

1．两个事件独立

给定事件 A 与 B，且 $P(A)>0$，$P(B)>0$，则

$$A 与 B 相互独立 \quad P(A|B)=P(A),$$

或
$$P(B|A)=P(B),$$

或
$$P(AB)=P(A)P(B).$$

2．三个事件独立

设 A, B, C 是三个事件，若满足等式

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

$$P(AC) = P(A)P(C),$$

$$P(BC) = P(B)P(C),$$

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

则称 A, B, C 相互独立，简称 A, B, C 独立．

要 点 范 例

一、随机事件的表示和运算

例 1.1 设 A, B, C 表示三个随机事件，试以 A, B, C 的运算表示下列事件：

(1) 仅 A 发生；

(2) A, B, C 都不发生；

(3) A, B, C 恰有一个发生．

解 (1) ABC；

(2) $\overline{AB}\,\overline{C}$；

(3) $ABC \cup \overline{AB}\,\overline{C} \cup ABC$．

注 ① 根据题目的含义按照事件交、并、逆的定义将复杂的事件用简单的运算予以表示．

② 在进行事件运算时，关于它们的运算次序作如下约定：先进行逆运算，其次进行积运算，最后进行并、差运算．

二、求古典概型的概率

例 1.2 从 $0, 1, \cdots, 9$ 十个数字中随机抽取一个数字，求取到奇数的概率．

解　令 W_i 表示取到的数字 $i(i=0,1,\cdots,9)$，$A=\{$取到奇数数字$\}$. $A=\{W_1, W_3, W_5, W_7, W_9\}$，即包含五个基本事件，又因为基本事件总数为十，从而

$$P(A)=\frac{5}{10}=\frac{1}{2}.$$

例 1.3　有 50 件产品，其中 45 件正品，5 件次品. 今从中任取 3 件，求下列事件的概率：

(1) $A=\{$恰取得 1 件次品$\}$；

(2) $B=\{$没有取得次品$\}$.

解　(1) 从 50 件产品中任意抽取 3 件产品，共有 C_{50}^3 种抽取方法，一种取法就是一个基本事件，则基本事件总数 $N=C_{50}^3$.

又 $A=\{$恰取得 1 件次品$\}=\{$取得 1 件次品，2 件正品$\}$，这一事件包含的基本事件个数可以这样计算：1 件次品从 5 件次品中取得，共有 C_5^1 种取法；2 件正品从 45 件正品中取得，共有 C_{45}^2 种取法. 因而共有 $C_5^1\times C_{45}^2$ 种取法，即 A 包含的基本事件数 $M=C_5^1\times C_{45}^2$. 这样

$$P(A)=\frac{C_5^1 C_{45}^2}{C_{50}^3}\approx 0.25.$$

(2) $B=\{$没有取得次品$\}=\{$取得 3 件正品$\}$，这一事件包含的基本事件数 $M=C_{45}^3$，所以

$$P(B)=\frac{C_{45}^3}{C_{50}^3}\approx 0.72.$$

注　这是典型的古典概型题，其一般模型为：有 N 个球，其中黑球为 M 个，白球为 $N-M$ 个，一次任取 n 个 $(n\leq N)$，在取得的 n 个球中恰有 m 个黑球的概率为

$$P=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

例 1.4　一批产品共 100 件，其中次品有 3 件，今从这批产品中接连抽取两次，每次抽取 1 件，考虑两种情形：

(1) 不放回抽样,第一次取 1 件不放回,第二次再抽 1 件;

(2) 放回抽样,第一次取 1 件检查后放回,第二次再抽 1 件 .

试分别就上述两种情况,求第一次抽到正品、第二次抽到次品的概率 .

解 (1) 采用不放回抽样,由于要考虑两件产品取出的顺序,接连两次抽取共有 A_{100}^2 种抽取方法,即 $N = A_{100}^2$.第一次抽到正品是从 97 个正品中取出,共有 97 种抽样方式;第二次取得次品是从 3 个次品中取出,共有 3 种抽取方式 .这样 $M = 97\times 3$.所求概率

$$P_1 = \frac{97\times 3}{A_{100}^2} \approx 0.0294 .$$

(2) 采用放回抽样时,第一次抽样有 100 种;因抽后又放回,所以第二次抽样方式还是 100 种 .这样,连续两次的抽样方式有 100^2 种,即 $N = 100^2$.在这种情况下, M 仍是 97×3 .因此所求事件的概率为

$$P_2 = \frac{97\times 3}{100^2} \approx 0.0291 .$$

注 ① 在概率论中,放回抽样与不放回抽样是两种不同的抽样方式 .在本例中,对两种抽样所求事件的概率数值不同,但相差无几,原因是产品总数很大而抽查产品很少 .

② 解这类题时,首先要判断基本事件是否具有等可能性,基本事件的总数是否有限,在肯定上述结论后应求得基本事件总数和事件 A 中所包含的基本事件数 .至于用排列还是组合应按题意分析基本事件的组成是否与排列次序有关,如无关则应用组合 .

例 1.5 袋中有 5 个白球和 3 个黑球,从袋中任取 2 个球,求:

(1) 取得同色 2 个球的概率;

(2) 取得的 2 个球至少有 1 个是白球的概率 .

解 (1) 从 8 个球中任取 2 个的取法总数为 $C_8^2 = 28$ 种 .取得

2 个同色球,可分为两种情况:2 个球皆为白色或 2 个球皆为黑色 . 这两种情况分别有 C_5^2, C_3^2 种取法,故取得 2 个同色球的取法为 $C_5^2 + C_3^2 = 13$ 种,故所求概率为

$$P_1 = \frac{13}{28} \approx 0.4643 .$$

此题也可以这样解:令

$A_1 = \{$取得的 2 个球皆为白色$\}$,

$A_2 = \{$取得的 2 个球皆为黑色$\}$,

$A = \{$取得的 2 个球同色$\}$.

A_1, A_2 互不相容,且 $A = A_1 \cup A_2$,而

$$P(A_1) = \frac{C_5^2}{C_8^2} = \frac{10}{28}, \quad P(A_2) = \frac{C_3^2}{C_8^2} = \frac{3}{28},$$

故 $$P_1 = P(A) = \frac{10}{28} + \frac{3}{28} = \frac{13}{28} \approx 0.4643 .$$

(2) 令 $A = \{$取得的 2 个球中至少有 1 个白球$\}$.则 $\overline{A} = \{$取得的 2 个球皆为黑球$\}$,故

$$P_2 = P(A) = 1 - P(\overline{A}) = 1 - \frac{C_3^2}{C_8^2} = 1 - \frac{3}{28}$$

$$= \frac{25}{28} \approx 0.8929 .$$

注 ① 本例告诉我们:当 $P(A)$不好求,但 $P(\overline{A})$好求时,可用 $P(A) = 1 - P(\overline{A})$来求 $P(A)$,这种技巧经常要用到 .

② 利用概率性质,将复杂事件分解为简单事件来处理,对复杂事件的概率运算是有益的 .

③ 利用概率运算的性质求事件的概率 .

例 1.6 设在一次试验中事件 A 发生的概率为 p,现进行 n 次独立试验,则 A 至少发生一次的概率为________ .

答 $1 - (1 - p)^n$.

分析 由 $P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - p(\overline{A_1})\, p(\overline{A_2}) \cdots p(\overline{A_n})$,

$$p(A_1) = p(A_2) = \cdots = p(A_n) = p,$$

$$p(\overline{A_1}) = p(\overline{A_2}) = \cdots = p(\overline{A_n}) = 1 - p(A_1) = 1 - p,$$

故
$$p(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - (1 - p)^n .$$

例 1.7 设 A, B, C 为三个事件, $P(A) = P(B) = P(C) = \frac{1}{4}$, $P(AB) = P(BC) = 0$, $P(AC) = \frac{1}{8}$, 则 $P(A \cup B \cup C) =$ ________ .

答 $\frac{5}{8}$.

分析 ABC AB, 故 $P(ABC) \leq P(AB) = 0$, 所以

$$P(ABC) = 0 .$$

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(AB) \\ &\quad - P(BC) - P(AC) + P(ABC) \\ &= \frac{5}{8} . \end{aligned}$$

例 1.8 袋中有黑、白球各 1 个, 一次次地从袋中随机地摸球, 每次摸球后(不管摸到的是白球, 或是黑球)都不把此球放回, 而是放进一只白球, 求第 n 次摸到的是白球的概率 .

解 原先袋中是 1 白 1 黑, 一次摸球后袋中的球可能是 1 黑 1 白或 2 白, 第一种情形的出现是因为摸到的白球, 第二种情形的出现是因为摸到的是黑球, 记 A 为所求的事件, B_n 为第 n 次摸到的是黑球, 则 $\overline{A} = B_1 B_2 \cdots B_n$.

$$\begin{aligned} P(\overline{A}) &= P(B_1 B_2 \cdots B_n) \\ &= P(B_1) P(B_2 \mid B_1) P(B_3 \mid B_2 B_1) \cdots P(B_n \mid B_{n-1} \cdots B_1) \\ &= {\frac{1}{2}}^{n}, \end{aligned}$$

所以
$$P(A) = 1 - \frac{1}{2^n} .$$

注 ① 对于并事件的概率通常利用概率的加法化为若干个交事件的概率的加减 .

② 对于交事件的概率一般利用概率乘法公式.当交事件为独立时可利用独立的性质.

③ 在某些情况下,利用概率乘法公式去解古典概型问题是比较简便的.

三、利用全概率公式计算复杂事件概率

例 1.9 某仓库有同样规格的产品 12 箱,其中有 6 箱、4 箱、2 箱分别是由甲、乙、丙 3 个厂生产的,3 个厂的次品率分别为 $\frac{1}{10}$,$\frac{1}{14}$,$\frac{1}{18}$.现从 12 箱中任取 1 箱,再从取得的 1 箱中任取 1 件产品,求取得的 1 件产品是次品的概率.

解 设 $A=\{$取得的 1 件产品是次品$\}$,

$B_1=\{$取得的 1 箱是甲厂的$\}$,

$B_2=\{$取得的 1 箱是乙厂的$\}$,

$B_3=\{$取得的 1 箱是丙厂的$\}$,

B_1,B_2,B_3 构成完备事件组.由题设得

$$P(B_1)=\frac{6}{12}=\frac{1}{2},\quad P(B_2)=\frac{4}{12}=\frac{1}{3},\quad P(B_3)=\frac{2}{12}=\frac{1}{6}.$$

且

$$P(A\mid B_1)=\frac{1}{10},\quad P(A\mid B_2)=\frac{1}{14},\quad P(A\mid B_3)=\frac{1}{18}.$$

由全概率公式得

$$P(A)=P(A\mid B_1)P(B_1)+P(A\mid B_2)P(B_2)+P(A\mid B_3)P(B_3)$$

$$=\frac{1}{2}\times\frac{1}{10}+\frac{1}{3}\times\frac{1}{14}+\frac{1}{6}\times\frac{1}{18}\approx 0.083.$$

注 ① 这是典型的全概率公式题型.A 是一个复杂事件,因取得的次品不知来自何厂.这里 $P(A\mid B_i)(i=1,2,3)$已知,要求 $P(A)$.

② 在计算复杂事件 A 的概率时,如果 A 事件的发生与多个事件 B_i 的发生有关,就应考虑用全概率公式.

③ 运用全概率公式的关键在于找一个完备事件组.然后分别按题意计算出在全概率公式中所需要的 $P(B_i)$和 $P(A|B_i)$.

例 1.10 一批晶体管元件,其中一等品占 95%,二等品占 4%,三等品占 1%,它们能工作 5 000 小时以上的概率分别为 90%,80%,70%.求任取一个元件能工作 5 000 小时以上的概率.

解 令 $B_i=\{$取到元件为 i 等品$\}(i=1,2,3)$, $A=\{$取到的元件能工作 5 000 小时以上$\}$,则

$$P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)$$
$$=95\%\cdot 90\%+4\%\cdot 80\%+1\%\cdot 70\%=0.894.$$

四、利用贝叶斯公式计算条件概率

例 1.11 电报信号由"·"与"－"组成,设发报台传递"·"与"－"之比为 3∶2.由于通讯系统存在干扰,可能引起失真.传递"·"时,失真的概率为 0.2(即发出"·"而收到"－");传送"－"时,失真的概率为 0.1(即发出"－"而收到"·").若收报台收到信号"·",求发报台确实发出"·"的概率.

解 设 $B_1=\{$发送"·"$\}$, $B_2=\{$发送"－"$\}$,
$A=\{$收到"·"$\}$.

B_1, B_2 构成完备事件组,且 $P(B_1)=0.6$, $P(B_2)=0.4$.又

$$P(A|B_1)=0.8,\quad P(A|B_2)=0.1.$$

由贝叶斯公式,所求概率为

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1)+P(B_2)P(A|B_2)}$$
$$=\frac{0.6\times 0.8}{0.6\times 0.8+0.4\times 0.1}\approx 0.923.$$

注 本例归结为贝叶斯公式题型.已知先验概率 $P(B_1)$,

$P(B_2)$,要求后验概率 $P(B_1|A)$.若收报台收到信号"－",发报台确实发出"－"的概率为何?请读者自己思考.

五、贝努利概型

例 1.12 设某人每次射击命中目标的概率为 p,试射击 n 次,求至少命中 2 次的概率.

解 这是贝努利概型,每次独立重复试验命中目标的概率为 p,按题意,设 n 次独立射击命中目标的次数为 m,所求概率为

$$
\begin{aligned}
P(m\geq 2) &= 1 - P_n(0) - P_n(1)\\
&= 1-(1-p)^n - C_n^1 p(1-p)^{n-1}\\
&= 1-(1-p)^n - np(1-p)^{n-1}.
\end{aligned}
$$

注 ① 本例若直接计算 $P(m\geq 2)=\sum\limits_{m=2}^{n} P_n(m)$ 则较繁琐.

② 要熟记概率公式: $P_n(m) = C_n^m p^m \cdot q^{n-m}$ $(m=0,1,2,\cdots,n)$.

③ 在 n 次独立重复试验中,事件 A 至少发生 r 次的概率为

$$P(m\geq r) = \sum_{m=r}^{n} P_n(m) = 1-\sum_{m=0}^{r-1} P_n(m).$$

特别 $$P(m\geq 1) = 1 - P_n(0) = 1-(1-p)^n.$$

习 题 选 解

(1－1) 设 A, B, C 表示 3 个随机事件,试以 A, B, C 的运算来表示下列事件:

(1) A, B, C 中恰好 1 个发生;

(2) A 不发生,而 B, C 中至少有 1 个发生;

(3) A, B, C 中至少有 2 个发生;

(4) A, B, C 中不多于 1 个发生.

解　(1)“A,B,C 中恰有 1 个发生”是一个复杂的事件，它可以分解为“A 发生，而 B,C 不发生”、“B 发生，而 A,C 不发生”、“C 发生，而 A,B 不发生”，它们可以分别表示为 $A\overline{B}\,\overline{C}$，$\overline{A}B\overline{C}$，$\overline{A}\,\overline{B}C$。这 3 个事件是互不相容的，它们的和事件即为所要表示的事件，即表示为：$A\overline{B}\,\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\,\overline{B}C$。

(2) 由于“B,C 中至少有 1 个发生”可表示为事件 $B\cup C$。所以“A 不发生，而 B,C 中至少有 1 个发生”可表示为 $\overline{A}\cap(B\cup C)$。

(3)“A,B,C 中至少有 2 个发生”这一事件可表示为 $AB\cup BC\cup CA$。

(4) 所述事件可以分解为“A 发生，B,C 不发生”、“B 发生，A,C 不发生”、“C 发生，A,B 不发生”、“A,B,C 都不发生”。它们分别表示为 $A\overline{B}\,\overline{C}$，$\overline{A}B\overline{C}$，$\overline{A}\,\overline{B}C$ 与 $\overline{A}\,\overline{B}\,\overline{C}$，它们的和事件 $A\overline{B}\,\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\,\overline{B}C\cup\overline{A}\,\overline{B}\,\overline{C}$ 即为所要表示的事件。

所述事件也可表示为

$$\overline{AB\cup BC\cup CA}.$$

(1-2)　随机点 x 落在区间 $[a,b]$ 上这一事件记做 $\{x\mid a\leq x\leq b\}$。设 $\Omega=\{x\mid -\infty<x<+\infty\}$，$A=\{x\mid 0\leq x<2\}$，$B=\{x\mid 1\leq x<3\}$。问下述运算分别表示什么事件？

(1) $A\cup B$;　(2) AB;　(3) $\overline{A}$;　(4) $A\overline{B}$.

解　(1) $A\cup B=\{x\mid 0\leq x<3\}$;

(2) $AB=\{x\mid 1\leq x<2\}$;

(3) $\overline{A}=\{x\mid -\infty<x<0, 2\leq x<+\infty\}$;

(4) $\overline{B}=\{x\mid -\infty<x<1$ 或 $x>3\}$.

故
$$A\overline{B}=\{x\mid 0\leq x<1\}.$$

注　计算 $A\overline{B}$ 时，应先求 $\overline{B}$。

(1-3)　若要击落飞机必须同时击毁 2 个发动机或击毁驾驶舱，记：

$$A_1=\{击毁第 1 个发动机\};$$

$$A_2=\{击毁第 2 个发动机\};$$

$$B=\{击毁驾驶舱\}.$$

试用 A_1, A_2 和 B 表示{飞机被击落}的事件.

解 依题意 A_1, A_2 同时发生或者 B 发生时飞机被击落,故

$$\{飞机被击落\}=A_1A_2\cup B.$$

(1-4) 已知 $A \quad B$, $P(A)=0.2$, $P(B)=0.3$, 求:

(1) $P(\overline{A})$, $P(\overline{B})$; (2) $P(A\cup B)$; (3) $P(AB)$;

(4) $P(B\overline{A})$; (5) $P(A-B)$.

解 (1) $P(\overline{A})=1-P(A)=0.8$, $P(\overline{B})=1-P(B)=0.7$.

(2) 因为 $A \quad B$, 所以 $P(A\cup B)=P(B)=0.3$.

(3) 因为 $A \quad B$, 所以 $P(AB)=P(A)=0.2$.

(4) $P(B\overline{A})=P(B-A)=P(B)-P(A)=0.3-0.2=0.1$.

(5) 因为 $A \quad B$, 所以

$$A-B=A\overline{B}=\text{ě},$$

$$P(A-B)=P(\text{ě})=0.$$

(1-5) 从 1,2,3,4,5 五个数码中,任取三个不同数码排成三位数,求:

(1) 所得三位数为偶数的概率;

(2) 所得三位数为奇数的概率.

解法一 (1) 三位数共有 A_5^3 种排法,其中末位数是 2 或 4 的即为偶数,十位百位上的数字从余下的 4 个数任取 2 个排列.因此,排得偶数的情形共有 $2A_4^2$ 种,故

$$P_1=\frac{2A_4^2}{A_5^3}=\frac{2\times4\times3}{5\times4\times3}=0.4.$$

(2) 同(1)作类似的分析,知

$$P_2=\frac{3A_4^2}{A_5^3}=\frac{3\times4\times3}{5\times4\times3}=0.6.$$

解法二 我们也可作如下分析:

{所得三位数是偶数} = {三位数的末位是偶数},又{所得三位数是奇数} = {三位数的末位是奇数},从而

$$P_1 = \frac{2}{5} = 0.4, \quad P_2 = \frac{3}{5} = 0.6.$$

注　显然,后一种分析使得计算简化了 .

(1 - 6)　一批零件共 100 个,次品率为 10% .每次从其中任取一个零件,共取 3 次,取出后不放回,求第 3 次才取得合格品的概率 .

解法一　第 3 次才取得合格品,意味着前 2 次取得的是次品,记 A_i 表示第 i 次取得合格品($i=1,2,3$) .所求概率为

$$\begin{aligned} P &= P(\overline{A_1}\,\overline{A_2}\,A_3) \\ &= P(\overline{A_1})P(\overline{A_2}\mid\overline{A_1})P(A_3\mid\overline{A_1}\,\overline{A_2}) \\ &= \frac{10}{100}\times\frac{9}{99}\times\frac{90}{98}\approx 0.008\,35 \end{aligned}$$

解法二　基本事件总数为 $A_{100}^3 = 100\times 99\times 98$, $A_1A_2A_3$ 包含的基本事件数为 $A_{10}^2\times 90$,故

$$P(A_1A_2A_3) = \frac{10\times 9\times 90}{100\times 99\times 98}\approx 0.008\,35.$$

(1 - 7)　10 个零件中有 3 个次品和 7 个合格品 .每次从其中任取 1 个零件,共取 3 次,取出后不放回 .求:

(1) 这 3 次都抽不到合格品的概率;

(2) 这 3 次中至少有 1 次抽到合格品的概率 .

解　(1) 记 A_i 表示第 i 次取得次品($i=1,2,3$),则

$$\begin{aligned} P(A_1A_2A_3) &= P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2) \\ &= \frac{3}{10}\times\frac{2}{9}\times\frac{1}{8} = \frac{6}{720}\approx 0.008\,3. \end{aligned}$$

(2)“3 次中至少有 1 次抽到合格品”的对立事件是“3 次都抽不到合格品”,故

$$P = 1 - P(A_1A_2A_3)\approx 0.991\,7.$$

(1 - 8)　设 10 件产品中有 4 件不合格,现从中连续抽取 2

次,每次 1 件,取出后不放回.求第 2 次取得合格品的概率.

解 设 $A=\{$第 1 次取到合格品$\}$,则 $A,\overline{A}$构成完备事件组.又设

$$B=\{\text{第 2 次取到合格品}\}.$$

由题设
$$P(A)=\frac{6}{10},\quad P(\overline{A})=\frac{4}{10},$$
$$P(B|A)=\frac{5}{9},\quad P(B|\overline{A})=\frac{6}{9},$$

由全概率公式

$$P(B)=\frac{6}{10}\times\frac{5}{9}+\frac{4}{10}\times\frac{6}{9}=\frac{6}{10}=0.6.$$

(1-9) 对以往数据分析结果表明,当机器调整良好时,产品的合格率为 90%;而当机器发生某一故障时,产品的合格率为 30%.机器开动时,机器调整良好的概率为 75%,试求已知某日首件产品是合格品时,机器调整良好的概率.

解 设 A 为事件"产品合格",B 为事件"机器调整良好".已知 $P(A|B)=0.9$, $P(A|\overline{B})=0.3$, $P(B)=0.75$, $P(\overline{B})=0.25$,所求概率为 $P(B|A)$.由贝叶斯公式

$$P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\overline{B})P(\overline{B})}$$
$$=\frac{0.9\times0.75}{0.9\times0.75+0.3\times0.25}=0.9.$$

(1-10) 一批产品共有 N 个,其中 M 个是次品.从这批产品中任取 1 个检查,记录其等级后,仍放回去,如此连续抽查 n 次.求:

(1) n 次都取得合格品的概率;

(2) n 次中至少有 1 次取得次品的概率.

解 因为取 1 个检查后再放回去,所以每次检查的次品率 $\frac{M}{N}$ 不变,由于是放回抽样,故每次抽取试验是相互独立的.

(1) 设 $A_i=\{$第 i 次取得合格品$\}(i=1,2,\cdots,n)$, A_1,

$A_2, \cdots, A_n$ 相互独立, $P(A_i) = \frac{N-M}{N} = 1 - \frac{M}{N}$.

又设 $B = \{ n$ 次都取得合格品$\}$,

则 $$P(B) = P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n)$$
$$= \left(1 - \frac{M}{N}\right)^n .$$

(2) $\overline{B} = \{ n$ 次中至少有 1 次取得次品$\}$,故

$$P(\overline{B}) = 1 - P(B) = 1 - \left(1 - \frac{M}{N}\right)^n .$$

(1－11) 证明:(1) 若 A 与 B 相互独立,则 A 与 $\overline{B}$ 相互独立.

(2) 设 $P(A) > 0$, $P(B) > 0$,若 A 与 B 相互独立,则 A 与 B 不可能互不相容;反之,若 A 与 B 互不相容,则 A 与 B 不可能相互独立.

证 (1) 因 $A\overline{B} = A - B$,故

$$P(A\overline{B}) = P(A - B) = P(A) - P(AB)$$
$$= P(A) - P(A)P(B) = P(A)(1 - P(B))$$
$$= P(A) \cdot P(\overline{B}),$$

故 A 与 $\overline{B}$ 相互独立.

(2) 若 A 与 B 相互独立,则

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

此时 A, B 不可能互不相容,因为若 A, B 互不相容,则有 $AB =$ ě ,必有 $P(AB) = P($ě$) = 0$,与上式矛盾.

若 A 与 B 互不相容, $AB =$ ě ,故

$$P(AB) = P(\text{ě}) = 0 .$$

此时 A, B 不可能相互独立,因为若 A, B 独立,则 $P(AB) = P(A)P(B) > 0$,与上式矛盾.

注 要重视互不相容与相互独立两个概念之间的区别."互不相容"指的是 A, B 不能同时发生,"相互独立"是指 A, B 发生互不影响.

(1－12) 加工某一零件共需经过 3 道工序.设第 1、2、3 道工

序的次品率分别是 2%,3%,5%.假定各道工序是互不影响的,求加工出来的零件的次品率.

解法一 设 A_i 表示第 i 道工序合格($i=1,2,3$),B 表示 3 道工序都合格,则 $\overline{B}=\{$至少有 1 道工序不合格$\}$.依题意要求 $P(\overline{B})$.因 A_1,A_2,A_3 相互独立,且

$$P(A_1)=0.98,\quad P(A_2)=0.97,\quad P(A_3)=0.95.$$

又 $B=A_1A_2A_3$,故由独立事件概率乘法公式,得

$$P(B)=P(A_1)P(A_2)P(A_3)$$

$$=0.98\times0.97\times0.95=0.903\ 07,$$

故

$$P(\overline{B})=1-0.903\ 07=0.096\ 93.$$

解法二

$$\overline{B}=\overline{A_1}\cup\overline{A_2}\cup\overline{A_3},$$

由对偶公式得

$$\overline{A_1}\cup\overline{A_2}\cup\overline{A_3}=\overline{A_1A_2A_3},$$

故

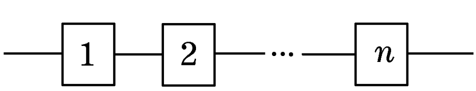
$$P(\overline{B})=P(\overline{A_1A_2A_3})=1-P(A_1A_2A_3)$$

$$=1-P(A_1)P(A_2)P(A_3)=0.096\ 93.$$

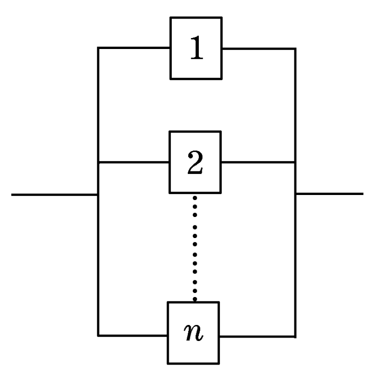
(1-13) 系统由 n 个元件连接而成.设第 i 个元件正常工作的概率为 $p_i(i=1,2,\cdots,n)$.求:

(1) 当 n 个元件按串联方式连接时(见图(a)),系统正常工作的概率;

(2) 当 n 个元件按并联方式连接时(见图(b)),系统正常工作的概率.



(a)



(b)

解　(1) 设元件正常工作是相互独立的，n 个元件串联时，要求整个系统正常工作，必须要求每个元件都正常工作.设

$$A_i=\{第\ i\ 个元件正常工作\}\quad (i=1,2,\cdots,n),$$

故
$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)=p_1p_2\cdots p_n.$$

(2) n 个元件并联时，要求整个系统正常工作，必须要求这 n 个元件至少有 1 个正常工作.其对立事件是 n 个元件都不正常工作.

$$\begin{aligned}P(A_1\cup A_2\cup\cdots\cup A_n)&=1-P(\overline{A_1}\,\overline{A_2}\cdots\overline{A_n})\\&=1-P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_n})\\&=1-(1-p_1)(1-p_2)\cdots(1-p_n).\end{aligned}$$

(1－14)　已知每枚地对空导弹击中来犯敌机的概率为0.96，问需要发射多少枚导弹才能保证至少有 1 枚导弹击中敌机的概率大于 0.999？

解　设需要发射 n 枚导弹才能保证至少有 1 枚导弹击中敌机的概率大于 0.999.这是贝努利概型.依题意所求概率为

$$P=1-P_n(0)=1-(0.04)^n>0.999,$$

故
$$n>\frac{\lg 0.001}{\lg 0.04}\approx 2.146,\qquad n=3.$$

(1－15)　一批产品中有 30% 的一级品，进行重复抽样调查，共取 5 个样品，求：

(1) 取出的 5 个样品中恰有 2 个一级品的概率；

(2) 取出的 5 个样品中至少有 2 个一级品的概率.

解　(1) 这是贝努利概型.$p=0.3$, $n=5$.所求概率为

$$P_5(2)=C_5^2\times 0.3^2\times 0.7^3=0.309.$$

(2) 依题意，所求概率为

$$\begin{aligned}P&=1-P_5(0)-P_5(1)\\&=1-0.7^5-C_5^1\times 0.3\times 0.7^4=0.472.\end{aligned}$$

(补 1－1)　设 A, B 是两个事件且 $P(A)=0.6$, $P(B)=0.7$.问：

(1) 在什么条件下 $P(AB)$ 取得最大值，最大值是多少？

(2) **在什么条件下** $P(AB)$**取得最小值,最小值是多少?**

解 (1) **当** $A \subset B$ **时**, $P(AB)$**取得最大值,此时**

$$P(AB) = P(A) = 0.6 .$$

(2) **当** $A \cup B = \Omega$ **时**, $P(AB)$**取得最小值**.

因为 $P(\Omega) = P(A \cup B) = P(A) + P(B) - P(AB) = 1$,

所以 $P(AB) = P(A) + P(B) - 1 = 0.6 + 0.7 - 1 = 0.3$.

(**补 1 - 2**) **已知** $P(A) = \dfrac{1}{4}$, $P(B|A) = \dfrac{1}{3}$, $P(A|B) = \dfrac{1}{2}$, **求** $P(A \cup B)$.

解 **因为**

$$P(A \cup B) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A) \cdot P(B|A). \qquad ①$$

又

$$P(A \cup B) = P(A) + P(B) - P(B) \cdot P(A|B). \qquad ②$$

由①,②式可得

$$P(B) = \frac{P(A) \cdot P(B|A)}{P(A|B)} = \frac{\dfrac{1}{4} \times \dfrac{1}{3}}{\dfrac{1}{2}} = \frac{1}{6}.$$

所以

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B|A) = \frac{1}{4} + \frac{1}{6} - \frac{1}{4} \times \frac{1}{3} = \frac{1}{3}.$$

(**补 1 - 3**) 3 **个人独立地去破译一份密码,已知各人能译出的概率分别为** $\dfrac{1}{5}, \dfrac{1}{3}, \dfrac{1}{4}$. **问** 3 **个人中至少有** 1 **个人能将此密码译出的概率是多少?**

解法一 **设** $A_i = \{$**第** i **人能译出密码**$\}$ $(i = 1, 2, 3)$. A_1, A_2, A_3 **相互独立**, $A = \{$**密码译出**$\}$, **则** $\overline{A} = \overline{A_1}\,\overline{A_2}\,\overline{A_3}$.

所以

$$\begin{aligned} P(A) &= 1 - P(\overline{A}) = 1 - P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}) \\ &= 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(\overline{A_3}) \\ &= 1 - \left(1 - \frac{1}{5}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{4}\right) = \frac{3}{5}. \end{aligned}$$

解法二 $P(A)=P(A_1\cup A_2\cup A_3)$

$$=P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3),$$

其中 $$P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3),$$

所以 $$P(A)=\frac{1}{5}+\frac{1}{3}+\frac{1}{4}-\frac{1}{5}\times\frac{1}{3}-\frac{1}{5}\times\frac{1}{4}-\frac{1}{3}\times\frac{1}{4}+\frac{1}{5}\times\frac{1}{3}\times\frac{1}{4}$$

$$=\frac{12}{60}+\frac{20}{60}+\frac{15}{60}-\frac{7}{60}-\frac{4}{60}=\frac{36}{60}=\frac{3}{5}.$$

试 题 选 解

[1－1] 填空题

1.同时掷两颗骰子,记录它们的点数之差的绝对值,则样本空间为________.

答 $\Omega=\{0,1,2,3,4,5\}$.

2.将一枚均匀硬币抛两次,用“1”表示出现正面,“0”表示出现反面,则样本空间是________.

答 $\Omega=\{(0,0),(0,1),(1,0),(1,1)\}$.

3.设 $A_n(n\geq 1)$为一事件列,则该事件列中至少有一个事件发生的事件可表达为________.

答 $A_1\cup A_2\cup\cdots\cup A_n$.

4.一口袋中装有 3 只红球,2 只黑球,今从中任意取出 2 只球,则这 2 只球恰为 1 红 1 黑的概率是________.

答 0.6.

因为所求概率 $P=\dfrac{C_3^1C_2^1}{C_5^2}=\dfrac{3}{5}=0.6$ **(古典概率)**.

5.设 $P(A)=\dfrac{1}{2}$, $P(B|A)=\dfrac{2}{5}$,**则** $P(AB)=$________.

答　$\frac{1}{5}$.

因为由概率乘法公式,得

$$P(AB) = P(B \mid A)P(A) = \frac{2}{5} \times \frac{1}{2} = \frac{1}{5}.$$

[1－2]　单选题

1. **设随机事件** A **与** B **互不相容,且** $P(A) > 0$, $P(B) > 0$,**则下列中正确的是**　(　　)

① $P(A) = 1 - P(B)$　　② $P(AB) = P(A)P(B)$

③ $P(A \cup B) = 1$　　④ $P(\overline{AB}) = 1$

答　选择④.因为 A 与 B 互不相容.故

$$AB = ě, \quad P(AB) = P(ě) = 0,$$

$$P(\overline{AB}) = 1 - P(AB) = 1 - P(ě) = 1.$$

2. **设** A, B **为随机事件**, $P(B) > 0$, $P(A \mid B) = 1$,**则必有**　(　　)

① $P(A \cup B) = P(A)$　　② $A \quad B$

③ $P(A) = P(B)$　　④ $P(AB) = P(A)$

答　选择①.

因为由公式 $P(A \cup B) = P(A) + P(B) - P(AB)$及 $P(AB) = P(B)P(A \mid B) = P(B)$立即可得.

3. **将** 2 **封信随机地投入** 4 **个邮筒中,则未向前面** 2 **个邮筒投信的概率为**　(　　)

① $\frac{2^2}{4^2}$　　② $\frac{C_2^1}{C_4^2}$　　③ $\frac{2!}{A_4^2}$　　④ $\frac{2!}{4!}$

答　选择①.

此处利用古典概型的计算公式可直接求得.

4. **某人连续向一目标射击,每次命中目标的概率为** $\frac{3}{4}$,**他连续射击直到命中为止,则射击次数为** 3 **的概率是**　(　　)

① $\left(\frac{3}{4}\right)^3$　　② $\left(\frac{3}{4}\right)^2 \times \frac{1}{4}$

③ $\left(\frac{1}{4}\right)^{2}\times\frac{3}{4}$　　　　④ $C_4^2\left(\frac{1}{4}\right)^{2}\frac{3}{4}$

答　选择③.

由于每次射击可视为互相独立的,设 A_i **表示第** i **次射中目标** $(i=1,2,3)$**.则所求概率为**

$$P(\overline{A_1}\ \overline{A_2}\ A_3)=P(\overline{A_1})P(\overline{A_2})P(A_3)=\left(\frac{1}{4}\right)^{2}\times\frac{3}{4}.$$

[1-3]　是非题

1.设 A,B,C **是三个事件且两两独立,则** A,B,C **相互独立.**　　(　　)

答　不正确.

因为根据 A,B,C **三个事件相互独立的定义,不仅要求三个事件中任取两个相互独立,而且还要求满足** $P(ABC)=P(A)P(B)P(C)$.

2.必然事件与任何事件独立.　　(　　)

答　正确.

设 A **为任何事件,** Ω **为必然事件,则有**

$$A\Omega=A,\quad \text{故}\ P(A\Omega)=P(A),\quad \text{而}\ P(\Omega)=1,$$

即有

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

所以 $P(A\Omega)=P(A)P(\Omega)$**,即** A **与** Ω **独立.**

3.设 A,B **为两个事件,若** $P(AB)=0$**,则** $P(A)=0$ **或** $P(B)=0$.　　(　　)

答　不正确.

反例:设 $\Omega=\{1,2,3,4,5,6,7\}$, $A=\{2,4,6\}$, $B=\{1,3,5,7\}$,

$$P(AB)=P(\text{ě})=0.$$

而

$$P(A)=\frac{3}{7}\neq 0,\quad P(B)=\frac{4}{7}\neq 0.$$

4.事件 A,B,C **中至少有两个发生可为** $AB\cup BC\cup AC$.

(　　)

答　正确.

由事件的和运算可得.

[1-4] 某城镇自行车共 10 000 辆,其牌号从 00 001 到 10 000,偶然遇到一辆自行车,问其牌号中有数字 8 的概率是多少?

解 设 $A=\{$牌号中有数字 8$\}$.

考虑所述事件的对立事件$\overline{A}$,即在千位、百位、十位、个位都不出现 8 的事件,其概率为 $\left(\frac{9}{10}\right)^4$,故

$$P(A)=1-P(\overline{A})=1-\left(\frac{9}{10}\right)^4.$$

[1-5] 某射击小组有 20 名射手,其中一级射手 4 人,二级射手 8 人,三级射手 7 人,四级射手 1 人,各级射手能通过选拔进入比赛的概率依次为 0.9,0.7,0.5,0.2.求任选一名射手能通过选拔进入比赛的概率.

分析 本题的解题关键在于找到一个完备事件组,然后利用全概率公式.

解 设 A 为所论的事件.$B_i=\{$所选射手为 i 级射手$\}(i=1,2,3,4)$,则 $B_i(i=1,2,3,4)$为一完备事件组,

$$A=AB_1\cup AB_2\cup AB_3\cup AB_4.$$

由于 $P(B_1)=\frac{4}{20}$, $P(B_2)=\frac{8}{20}$, $P(B_3)=\frac{7}{20}$, $P(B_4)=\frac{1}{20}$,

因此,

$$P(A)=\sum_{i=1}^{4}P(A\mid B_i)P(B_i)$$

$$=0.9\times\frac{4}{20}+0.7\times\frac{8}{20}+0.5\times\frac{7}{20}+0.2\times\frac{1}{20}=0.645.$$

[1-6] 两台机床加工同样的零件,第一台的废品率为0.03,第二台的废品率为 0.02.加工出来的零件放在一起,且已知第一台加工的零件比第二台加工的多一倍.求任意取出的零件是合格品的概率.又若任意取出的零件经检查是废品,求它是第二台机床

加工的概率.

解　设 $A=\{$任取零件为合格品$\}$,

$B_1=\{$任取零件来自第一台机床加工的零件$\}$,

$B_2=\{$任取零件来自第二台机床加工的零件$\}$,

由全概率公式,有

$$P(\overline{A})=P(B_1)P(\overline{A}\mid B_1)+P(B_2)P(\overline{A}\mid B_2)$$
$$=\frac{2}{3}\times 0.03+\frac{1}{3}\times 0.02=0.027,$$

故
$$P(A)=1-0.027=0.973.$$

又
$$P(B_2\mid\overline{A})=\frac{P(B_2)P(\overline{A}\mid B_2)}{P(\overline{A})}=\frac{\frac{1}{3}\times 0.02}{\frac{1}{3}\times 0.08}=\frac{1}{4}.$$

[1-7]　设有甲、乙两个相同的口袋,甲袋中装有 3 个白球、5 个黑球,乙袋中装有 4 个白球、2 个黑球.某人从甲袋中任取 1 个球投入乙袋,然后再从乙袋中任取 1 个球.问这球是白球的概率为多少?

解　设 $A=\{$从乙袋中任取 1 个球是白球$\}$;

$B_1=\{$从甲袋中任取 1 个球是白球$\}$;

$B_2=\{$从甲袋中任取 1 个球是黑球$\}$.

由全概率公式,得

$$P(A)=P(B_1)P(A\mid B_1)+P(B_2)P(A\mid B_2)$$
$$=\frac{C_3^1}{C_8^1}\cdot\frac{C_5^1}{C_7^1}+\frac{C_5^1}{C_8^1}\cdot\frac{C_4^1}{C_7^1}=\frac{35}{56}=\frac{5}{8}.$$

[1-8]　证明题

设 A, B 为两个随机事件,$0<P(B)<1$,且 $P(A\mid B)=P(A\mid\overline{B})$,证明事件 A 与 B 相互独立.

提示　要证事件 A 与 B 相互独立,按定义,即要证明

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

证法一　由题设及条件概率定义得

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

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

由以上两式可得

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

故 A 与 B 相互独立.

证法二　由全概率公式得

$$\begin{aligned} P(A) &= P(B)P(A|B) + P(\overline{B})P(A|\overline{B}) \\ &= [P(B) + P(\overline{B})]P(A|B) \quad (\text{由题设}) \\ &= P(A|B), \end{aligned}$$

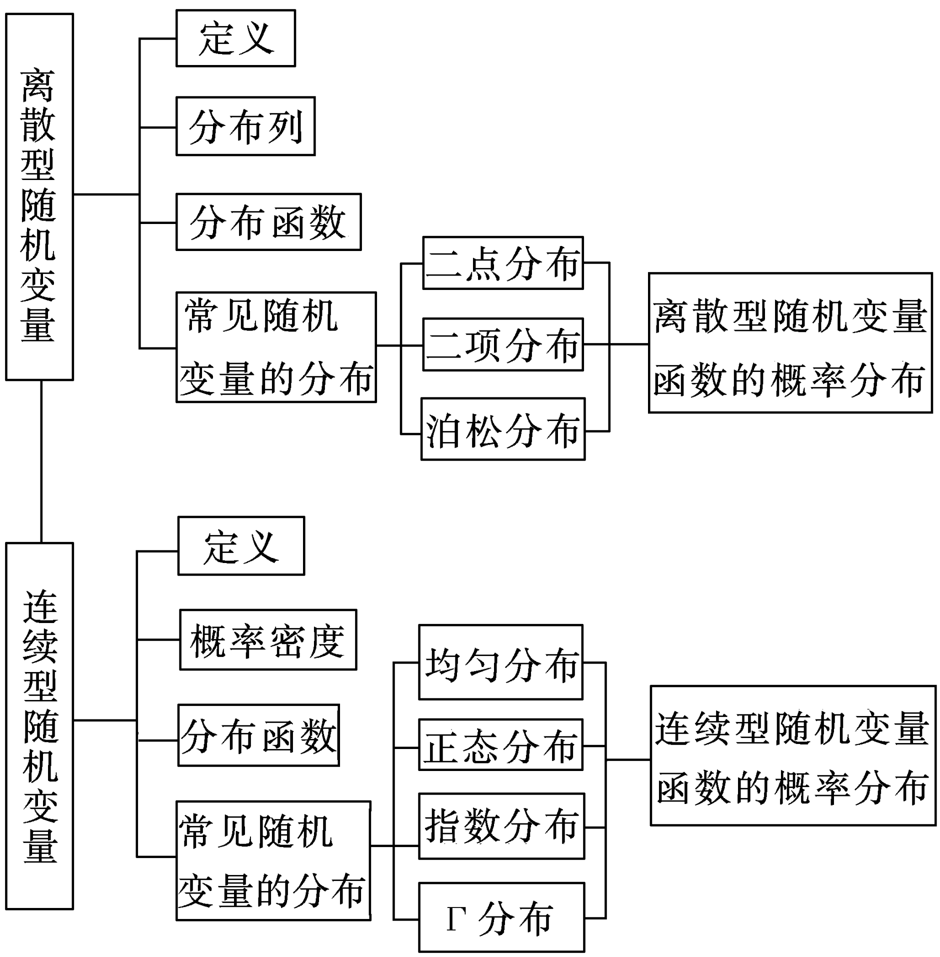
则 $$P(AB) = P(B)P(A|B) = P(A)P(B),$$

故 A 与 B 相互独立.

第二章　随机变量与概率分布

内 容 提 要

一、知识脉络



二、知识要点

(一) 随机变量

设 E 是一随机试验, Ω 是其样本空间.若对 $\forall\, \omega \in \Omega$ 都有一实数 $X(\omega)$与之对应,则称 $X = X(\omega)$为随机变量.

(二) 离散型随机变量

1.分布列

设离散型随机变量 X 的所有可能取值为 $x_k(k=1,2,\cdots)$,则 X 取各可能值的概率

$$P\{X = x_k\} = p_k, \quad k = 1,2,\cdots,$$

称为 X 的分布列.

2.分布列的性质

分布列有如下性质:

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

(2) $\sum\limits_k p_k = 1$.

3.重要分布

几种离散型随机变量的重要分布:

(1) (0 - 1 分布) 设随机变量 X 的分布列为

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

称 X 服从两点分布.

(2) 二项分布 设随机变量 X 的分布列为

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}, \quad k=0,1,2,\cdots,n,$$

称 X 服从参数为 n, p 的二项分布,记为 $X \sim B(n,p)$.

(3) 泊松分布 设随机变量 X 的分布列为

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

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

(三) 连续型随机变量

1.概率密度

连续型随机变量 X 所对应的理论分布曲线函数 $f(x)$ 称为 X 的概率密度函数,简称为概率密度.

2.概率密度的性质

设 $f(x)$ 是概率密度函数,则

(1) $f(x)\geq 0$;

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

3.重要分布

几种对连续型随机变量的重要分布.

(1) 均匀分布　若随机变量 X 具有概率密度

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

称 X 在 $[a,b]$ 上服从均匀分布,记为 $X\sim U(a,b)$.

(2) 正态分布　若随机变量 X 具有概率密度

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty,$$

称 X 服从参数为 $\mu,\sigma>0$ 的正态分布,记为 $X\sim N(\mu,\sigma^2)$.

注　① 若 $\mu=0,\sigma=1$,称 $N(0,1)$ 为标准正态分布.

② 若 $X\sim N(\mu,\sigma^2)$,则 $\dfrac{X-\mu}{\sigma}\sim N(0,1)$,且有

$$P\{x_1\leq X\leq x_2\}=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right),$$

其中

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t.$$

③ $\Phi(-x)=1-\Phi(x)$, $\Phi(0)=\dfrac{1}{2}$.

(3) 指数分布　若随机变量 X 具有概率密度

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

称 X 服从参数为 λ 的指数分布.

(四) 分布函数

1. 定义

设 X 是一随机变量,对任何实数 x,令

$$F(x)=P\{X\leq x\},\quad x\in(-\infty,+\infty),$$

称 $F(x)$为随机变量 X 的分布函数.

若 X 是离散型随机变量,则

$$F(x)=\sum_{x_i\leq x}p_i,$$

其中 $P\{X=x_i\}=p_i$ 为分布列;若 x 是连续型随机变量,则

$$F(x)=\int_{-\infty}^{x}f(t)\,\mathrm{d}t,$$

其中 $f(x)$为概率密度.

2. 性质

分布函数 $F(x)$具有以下基本性质:

(1) $F(x)$是 x 的不减函数,即若 $x_1<x_2$,则有

$$F(x_1)\leq F(x_2);$$

(2) $0\leq F(x)\leq 1,\ -\infty<x<+\infty;$

(3) $F(+\infty)=\lim\limits_{x\to+\infty}F(x)=\int_{-\infty}^{+\infty}f(x)\,\mathrm{d}x=1;$

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0.$$

利用分布函数的定义和性质可得到

$$P\{x_1<X\leq x_2\}=F(x_2)-F(x_1),\quad x_1<x_2.$$

(五) 随机变量函数的分布

1. 离散型情形

若随机变量 X 的分布列为

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p	p_1	p_2	$\cdots$	p_n	$\cdots$

则 $Y = g(x)$的分布律为

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_n)$	$\cdots$
p	p_1	p_2	$\cdots$	p_n	$\cdots$

若某些 $g(x_i)$相等,则应并作一项,概率相加 .

2 . 连续型情形

已知连续型随机变量 X 的概率密度为 $f(x)$, $y = g(x)$,求 y 的概率密度一般采用以下方法:先求出 y 的分布函数 $F_Y(y)$;再求 y 的概率密度 $f_Y(y)$, $f_Y(y) = \dfrac{\mathrm{d}}{\mathrm{d}y}F_Y(y)$.

若 X 具有概率密度 $f(x)$,若 $y = g(x)$是严格单调函数且可导,则 $Y = g(X)$是一个连续型随机变量,其概率密度为

$$f_Y(y) = \begin{cases} f[h(y)]\,|h'(y)|, & \alpha < y < \beta; \\ 0, & \text{其他} . \end{cases}$$

其中 $h(y)$是 $g(x)$的反函数,(α, β)是 y 的取值范围 .

$$\alpha = \min\{g(-\infty), g(+\infty)\},$$

$$\beta = \max\{g(-\infty), g(+\infty)\} .$$

要 点 范 例

一、求离散型随机变量的概率分布

例 2 .1 连续两次对目标进行射击,设每次击中目标的概率都为 0 .4, X 为击中目标的次数,求 X 的分布列 .

解 由题设, $X \sim B(2, 0.4)$.故

$$P\{X=0\}=0.6^2=0.36,$$
$$P\{X=1\}=2\times 0.4\times 0.6=0.48,$$
$$P\{X=2\}=0.4^2=0.16.$$

得分布列

X	0	1	2
p	0.36	0.48	0.16

例 2.2 袋中有 2 个白球和 3 个黑球，每次从其中任取 1 个球，直至取得白球为止. 记随机变量 X 为直到取得白球时的取球次数，在下列两种情形下，分别求 X 的分布列：

(1) 每次取出的黑球不再放回去；

(2) 每次取出的黑球仍放回去.

解 (1) 因为每次取出的黑球不再放回去，所以 X 的所有可能取值是 1, 2, 3, 4. 由概率乘法公式易知：

$$p_1=P\{X=1\}=\frac{2}{5}=0.4,$$
$$p_2=P\{X=2\}=\frac{3}{5}\times\frac{2}{4}=0.3,$$
$$p_3=P\{X=3\}=\frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}=0.2,$$
$$p_4=P\{X=4\}=\frac{3}{5}\times\frac{2}{4}\times\frac{2}{3}\times\frac{1}{2}=0.1.$$

因此，X 的分布列为

X	1	2	3	4
p	0.4	0.3	0.2	0.1

(2) 因为每次取出的黑球仍放回去，所以 X 的所有可能取值是一切正整数：$1, 2, \cdots, n, \cdots$. 且每次取球是相互独立的. 由独立事件的概率乘法公式，得：

$$p_1 = P\{X=1\} = \frac{2}{5},$$

$$p_2 = P\{X=2\} = \frac{3}{5}\times\frac{2}{5} = \frac{6}{25},$$

$$p_3 = P\{X=3\} = \left(\frac{3}{5}\right)^2\times\frac{2}{5} = \frac{18}{125},$$

一般地，

$$p_n = P\{X=n\} = \left(\frac{3}{5}\right)^{n-1}\times\frac{2}{5}$$

$$= 0.4\times(0.6)^{n-1},\quad n=1,2,\cdots.$$

因此，X 的分布列为

X	1	2	$\cdots$	n	$\cdots$
p	0.4	0.4×0.6	$\cdots$	$0.4\times(0.6)^{n-1}$	$\cdots$

注　在求离散型随机变量的概率分布时要注意两点：

① 考虑随机变量所有可能的取值；

② $\sum_i p_i = 1$.

二、求连续型随机变量的概率密度及随机事件的概率

例 2.3　设随机变量 X 的概率密度为

$$f(x) = \begin{cases} \dfrac{C}{\sqrt{1-x^2}}, & \text{当}|x|<1; \\ 0, & \text{当}|x|\geq 1. \end{cases}$$

其中 C 为待定常数.

(1) 确定 C 值；

(2) 求 X 落入 $\left(-3,\frac{1}{2}\right)$ 内的概率.

解　(1) 由概率密度性质得

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \int_{-1}^{1}\frac{C}{\sqrt{1-x^2}}\,\mathrm{d}x = C\arcsin x\Big|_{-1}^{1}$$

$$= C\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = C\pi = 1,$$

因此，
$$C = \frac{1}{\pi}.$$

(2) 由于 $f(x)$ 是分段表示的，求 $P\left\{-3 < X < \frac{1}{2}\right\}$ 需分段积分.

$$P\left\{-3 < X < \frac{1}{2}\right\} = \int_{-3}^{-1} 0\,\mathrm{d}x + \int_{-1}^{\frac{1}{2}} \frac{1}{\pi\sqrt{1 - x^2}}\,\mathrm{d}x$$

$$= \frac{1}{\pi}\arcsin x\Big|_{-1}^{\frac{1}{2}} = \frac{1}{\pi}\left(\frac{\pi}{3} - \left(-\frac{\pi}{2}\right)\right) = \frac{5}{6}.$$

例 2.4 公共汽车站每隔 5 分钟有一辆汽车通过，又设乘客在 5 分钟内任一时刻到达汽车站是等可能的，求乘客候车时间不超过 3 分钟的概率.

解 设 X 是 $[0,5]$ 之间乘客到达的时刻，依题意，X 服从均匀分布 $U(0,5)$. 概率密度为 $f(x) = \begin{cases} \frac{1}{5}, & 0 \leq x \leq 5; \\ 0, & \text{其他}. \end{cases}$ 所求概率为

$$P\{X \leq 3\} = \int_0^3 \frac{1}{5}\,\mathrm{d}x = \frac{3}{5}.$$

例 2.5 已知某种类型的电子管的寿命 X(以小时计)服从指数分布，其概率密度为

$$f(x) = \begin{cases} \frac{1}{1\,000}\mathrm{e}^{-\frac{x}{1\,000}}, & x > 0; \\ 0, & \text{其他}. \end{cases}$$

一台仪器中装有 5 只此类型电子管，任一只损坏时仪器便不能正常工作. 求仪器正常工作 1 000 小时以上的概率.

解 首先求一只电子管工作 1 000 小时以上的概率，

$$p_1 = P\{X > 1\,000\} = \int_{1\,000}^{+\infty} \frac{1}{1\,000}\mathrm{e}^{-\frac{x}{1\,000}}\,\mathrm{d}x = \mathrm{e}^{-1} \approx 0.367\,9.$$

只有当 5 只电子管皆工作在 1 000 小时以上，仪器才能工作 1 000

小时以上.又“每只电子管工作 1 000 小时以上”是相互独立的,所以所求概率为

$$p_5 = p_1^5 = e^{-5} \approx 0.006\,73.$$

三、求随机变量的分布函数

例 2.6 接连进行两次射击,设每次击中目标的概率为 0.4,设 X 为击中目标的次数,求 X 的分布函数.

解 由例 2.1 可知 X 的分布列为

X	0	1	2
p	0.36	0.48	0.16

当 $x<0$ 时,

$$F(x) = P\{X \leq x\} = 0;$$

当 $0 \leq x < 1$ 时,

$$F(x) = P\{X = 0\} = 0.36;$$

当 $1 \leq x < 2$ 时,

$$F(x) = P\{X = 0\} + P\{X = 1\} = 0.84;$$

当 $x \geq 2$ 时,

$$F(x) = P\{X = 0\} + P\{X = 1\} + P\{X = 2\} = 1.$$

故

$$F(x) = \begin{cases} 0, & x < 0; \\ 0.36, & 0 \leq x < 1; \\ 0.84, & 1 \leq x < 2; \\ 1, & x \geq 2. \end{cases}$$

注 求离散随机变量 X 的分布函数 $F(x)$,要按 $F(x)$ 的定义计算,它为 $(-\infty, x]$ 中包含的可能值 x_i 所对应的概率之和.

例 2.7 设离散随机变量 X 的分布列为

X	0	1	2
p	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$

求:(1) X **的分布函数** $F(x)$;

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

解 (1) **当** $x<0$ **时**, $F(x)=P\{X\leq x\}=0$;

当 $0\leq x<1$ **时**,

$$F(x)=P\{X\leq x\}=P\{X=0\}=\frac{1}{3};$$

当 $1\leq x<2$ **时**,

$$F(x)=P\{X\leq x\}=P\{X=0\}+P\{X=1\}$$

$$=\frac{1}{3}+\frac{1}{6}=\frac{1}{2};$$

当 $x\geq 2$ **时**,

$$F(x)=P\{X\leq 2\}=P\{X=0\}+P\{X=1\}+P\{X=2\}=1\ .$$

故

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

(2) $P\ X\leq\frac{1}{2}\ =F\ \frac{1}{2}\ =\frac{1}{3}\ .$

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

或者,由 $1,\frac{3}{2}$ **中未包含** X **的可能值,得**

$$P\ 1<X\leq\frac{3}{2}\ =P(\text{ě}\ )=0\ .$$

(3) $P\ 1\leq X\leq\frac{3}{2}\ =P\{X=1\}=\frac{1}{6}\ .$

例 2 .8 设随机变量 X **具有概率密度**

$$f(x)=\frac{a}{1+x^2}\quad(-\infty<x<+\infty)\ .$$

(1) **确定系数** a;

(2) 求分布函数;

(3) 计算 $P\{-1<X<1\}$.

解 (1) 利用概率密度性质

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=\int_{-\infty}^{+\infty}\frac{a}{1+x^2}\mathrm{d}x = a\arctan x\Big|_{-\infty}^{+\infty} = a\pi = 1,$$

得
$$a=\frac{1}{\pi}.$$

(2) $F(x)=\int_{-\infty}^{x}\frac{1}{\pi(1+t^2)}\mathrm{d}t=\frac{1}{\pi}\arctan x+\frac{1}{2}$.

(3) $P\{-1<X<1\}=P\{-1<X\leq 1\}=F(1)-F(-1)$

$$=\frac{1}{\pi}[\arctan 1-\arctan(-1)]=\frac{1}{2}.$$

例 2.9 某机器生产的螺栓长度 X(单位:cm)服从正态分布 $N(10.05, 0.06^2)$,规定长度在范围 10.05 ± 0.12 内为合格,求螺栓不合格的概率.

解 根据题设 $X\sim N(10.05, 0.06^2)$,化为标准正态分布计算.

螺栓合格的概率可表示为

$P\{10.05-0.12\leq X\leq 10.05+0.12\}$

$=\Phi\left(\frac{10.15+0.12-10.05}{0.06}\right)-\Phi\left(\frac{10.05-0.12-10.05}{0.06}\right)$

$=\Phi(2)-\Phi(-2)=2\Phi(2)-1=0.9545$ (查表求 $\Phi(2)$).

故螺栓不合格的概率为

$$p=1-0.9545=0.0455.$$

注 对于有关正态分布 $N(\mu,\sigma^2)$问题,熟练掌握以下公式将为计算带来很大方便:

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

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

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

四、求随机变量 X 的函数的分布

例 2.10 设随机变量 X 的分布列为

X	-2	-1	0	1	2
p	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{11}{30}$

试求:(1) $Y=2X+1$ 的分布列;

(2) 求 $Z=X^2+1$ 的分布列.

解 将 Y、Z 的可能取值及对应的概率列表如下

X	-2	-1	0	1	2
Y	-3	-1	1	3	5
Z	5	2	1	2	5
p	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{11}{30}$

整理后得

(1)

$Y=2X+1$	-3	-1	1	3	5
p	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{11}{30}$

(2)

$Z=X^2+1$	1	2	5
p	$\frac{1}{5}$	$\frac{7}{30}$	$\frac{17}{30}$

例 2.11 已知 X 的分布列为

X	1	2	3	$\cdots$	n	$\cdots$
p	$\frac{1}{2}$	$\left(\frac{1}{2}\right)^2$	$\left(\frac{1}{2}\right)^3$	$\cdots$	$\left(\frac{1}{2}\right)^n$	$\cdots$

求 $Y=\sin\frac{\pi}{2}X$ 的分布列.

解 因为

$$\sin\frac{n\pi}{2}=\begin{cases}-1, & \text{当 } n=4k-1;\\ 0, & \text{当 } n=2k; \qquad (k\in\mathbf{N}),\\ 1, & \text{当 } n=4k-3.\end{cases}$$

所以 $Y=\sin\frac{\pi}{2}X$ **只有 3 个可能值**: $-1, 0, 1$.

$$P\{Y=-1\}=\sum_{k=1}^{\infty}P\{X=4k-1\}=\sum_{k=1}^{\infty}\left(\frac{1}{2}\right)^{4k-1}$$

$$=\frac{\frac{1}{2^3}}{1-\frac{1}{16}}=\frac{2}{15};$$

$$P\{Y=0\}=\sum_{k=1}^{\infty}P\{X=2k\}=\sum_{k=1}^{\infty}\left(\frac{1}{2}\right)^{2k}$$

$$=\frac{\frac{1}{2^2}}{1-\frac{1}{4}}=\frac{1}{3};$$

$$P\{Y=1\}=\sum_{k=1}^{\infty}P\{X=4k-3\}=\sum_{k=1}^{\infty}\left(\frac{1}{2}\right)^{4k-3}$$

$$=\frac{\frac{1}{2}}{1-\frac{1}{16}}=\frac{8}{15}.$$

所以 Y **的分布列为**

Y	-1	0	1
p	$\frac{2}{15}$	$\frac{1}{3}$	$\frac{8}{15}$

例 2.12　求 $Y=kX+b(k\neq 0)$**的概率密度**,**设**:

(1) X **具有概率密度** $f_X(x)$;

(2) X **服从正态分布** $N(\mu,\sigma^2)$.

解　(1) $y=kx+b$ **则** $x=\frac{y-b}{k}$, $\frac{\mathrm{d}x}{\mathrm{d}y}=\frac{1}{k}$, **所以**

$$f_Y(y)=\frac{1}{|k|}f_X\left(\frac{y-b}{k}\right),\quad -\infty<y<+\infty.$$

(2) 当 $X\sim N(\mu,\sigma^2)$时,因为

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

由(1)有

$$f_Y(y)=\frac{1}{\sqrt{2\pi}|k|\sigma}\mathrm{e}^{-\frac{(y-k\mu-b)^2}{2k^2\sigma^2}},\quad -\infty<y<+\infty.$$

例 2.13 对圆片直径进行测量,测量值 X 服从均匀分布 $U(5,6)$,求圆面积 Y 的概率密度.

解 $Y=\frac{1}{4}\pi X^2$, X 的概率密度为

$$f_X(x)=\begin{cases}1, & 5\leq x\leq 6;\\ 0, & \text{其他}.\end{cases}$$

当$\frac{25}{4}\pi\leq y\leq 9\pi$时, $y=\frac{\pi}{4}x^2$ 单调增加, $x=\sqrt{\frac{4y}{\pi}}=2\sqrt{\frac{y}{\pi}}=h(y)$, $h'(y)=\frac{1}{\sqrt{\pi y}}$,故当$\frac{25}{4}\pi\leq y\leq 9\pi$时,

$$f_Y(y)=1\times\frac{1}{\sqrt{\pi y}}=\frac{1}{\sqrt{\pi y}}.$$

而当 y 取其他值时, $f_Y(y)=0$,故

$$f_Y(y)=\begin{cases}\frac{1}{\sqrt{\pi y}}, & \frac{25}{4}\pi\leq y\leq 9\pi;\\ 0, & \text{其他}.\end{cases}$$

注 ① 对离散型随机变量 X 的函数 $Y=g(X)$的概率分布只须确定 Y 的值域,并将具有相同 Y 值的 X 取值点的概率相加.

② 对连续型随机变量 X 的函数 $Y=g(X)$而言,若 Y 是连续型的,首先要确定 Y 的值域,然后求 Y 的分布函数 $F_Y(y)=P\{g(X)\leq y\}$,最后对 $F_Y(y)$求导得 Y 的概率密度 $f_Y(y)$.当 $Y=g(X)$为单调函数时,可利用 $f_Y(y)=f_X[h(y)]|h'(y)|$,其中

$x = h(y)$是 $y = g(x)$的反函数.

习 题 选 解

(2-1) 盒中有 12 只晶体管,其中有 2 只次品,10 只正品.现从盒中任取 3 只.求取出的 3 只所含次品数 X 的分布列.

解 因为在抽样中的次品数可以是 0,1,2,故 X 所有可能取值为 0,1,2.故有

$$P\{X=0\} = \frac{C_{10}^3}{C_{12}^3} = \frac{6}{11} \approx 0.545;$$

$$P\{X=1\} = \frac{C_{10}^2 C_2^1}{C_{12}^3} = \frac{9}{22} \approx 0.409;$$

$$P\{X=2\} = \frac{C_{10}^1 C_2^2}{C_{12}^3} = \frac{10\times 1}{220} \approx 0.046.$$

由此得 X 的分布列为

X	0	1	2
p	0.545	0.409	0.046

(2-2) 抛掷硬币 10 次,写出国徽向上次数 X 的分布列;并求国徽向上次数不小于 3 的概率.

解 依题意,$X\sim B(10, 0.5)$.故

$$P\{X=i\} = C_{10}^i 0.5^i 0.5^{10-i}, \quad i = 0,1,\cdots,10.$$

$$\begin{aligned} P\{X\geq 3\} &= 1 - P\{X<3\} \\ &= 1 - P\{X=0\} - P\{X=1\} - P\{X=2\} \\ &= 1 - [0.5^{10} + 10\times 0.5^{10} + 45\times 0.5^{10}] \\ &= 1 - 56\times 0.5^{10} \approx 0.945\,3. \end{aligned}$$

(2-3) 设一批产品共 2 000 个,其中有 40 个次品.随机抽取 100 个样品,求样品中次品数 X 的分布列,分别按下列抽样方

式求:

(1) 不放回抽样;

(2) 放回抽样.

解　(1) 不放回抽样相当于一次性取样,因此有

$$P\{X=i\}=\frac{C_{40}^{i}C_{1\,960}^{100-i}}{C_{2\,000}^{100}},\quad i=0,1,\cdots,40.$$

(2) 放回抽样相当于独立重复试验.抽取次品的概率

$$p=0.02,\quad X\sim B(100,0.02),$$

有 $$P\{X=i\}=C_{100}^{i}\times 0.02^{i}\times 0.98^{100-i},\quad i=0,1,\cdots,100.$$

(2-4)　设事件 A 在每次试验中发生的概率为 0.3.当 A 发生不少于 3 次时,指示灯发出信号,求:

(1) 进行 5 次独立试验,指示灯发出信号的概率;

(2) 进行 7 次独立试验,指示灯发出信号的概率.

解　(1) 此题为贝努利概型.每次试验 A 发生的概率 $p=0.3$,设 5 次独立试验中 A 发生的次数为 X, $X\sim B(5,0.3)$,要求 $P\{X\geq 3\}$.

$$P\{X\geq 3\}=C_5^3\times 0.3^3\times 0.7^2+C_5^4\times 0.3^4\times 0.7+0.3^5$$
$$\approx 0.132\,3+0.028\,35+0.002\,43\approx 0.163.$$

(2) 此题也为贝努利概型.$p=0.3$, $n=7$, $X\sim B(7,0.3)$.

$$P\{X\geq 3\}=1-P\{X\leq 2\}$$
$$=1-[0.7^7+C_7^1\times 0.3\times 0.7^6+C_7^2\times 0.3^2\times 0.7^5]$$
$$\approx 0.353.$$

(2-5)　一批产品中有 15% 的次品.进行独立重复抽样检查,共抽取 20 个样品,问取出的 20 个样品中最大可能的次品数是多少?并求其概率.

解　$n=20$, $p=0.15$.当 $i=[(n+1)p]=[21\times 0.15]$[①] $=[3.15]=3$ 时, $P_i=C_n^i p^i q^{n-i}$ 取得最大值($p=0.15$, $q=1-p=$

① [·]为取整函数符号.

0.85).

$$P_3 = C_{20}^2 \times 0.15^3 \times 0.85^{17}$$
$$= 1\,140 \times 0.003\,375 \times 0.063\,113 = 0.242\,8.$$

(2-6)　一电话交换台每分钟的呼唤次数服从参数为 4 的泊松分布.求:

(1) 每分钟恰有 6 次呼唤的概率;

(2) 每分钟呼唤次数不超过 10 的概率.

解　(1) **设** X **为每分钟呼唤次数,则** $X \sim P(4)$,**故**

$$P\{X = 6\} = \frac{4^6}{6!}e^{-4} \approx 0.104\,2.$$

(2) $P\{X \leq 10\} = 1 - P\{X \geq 11\} = 1 - \sum_{i=11}^{\infty} \frac{4^i}{i!}e^{-4}.$

查附表 2,**得**

$$P\{X \geq 11\} = 0.002\,84.$$

故

$$P\{X \leq 10\} = 1 - 0.002\,84 = 0.997\,16.$$

(2-7)　有一汽车站有大量汽车通过,设每辆汽车在一天某段时间出事故的概率为 0.000 1.在某天该段时间内有 1 000 辆汽车通过,求事故次数不少于 2 的概率.

解　$p = 0.000\,1$, $n = 1\,000$, **设事故次数为** X, **则** $X \sim B(1\,000, 0.000\,1)$, **因** n **较大**, p **很小**, $np = 0.1$, X **近似服从泊松分布** $P(0.1)$, **故**

$$P\{X \geq 2\} = \sum_{i=2}^{\infty} \frac{0.1^i}{i!}e^{-0.1}$$
$$= 1 - P\{X = 0\} - P(X = 1)$$
$$= 1 - e^{-0.1} - 0.1e^{-0.1}$$
$$= 1 - 1.1e^{-0.1} \approx 0.004\,68.$$

(2-8)　函数 $\sin x$ **是否为随机变量** X **的概率密度,如果** X **的可能值充满区间:**

(1) $\left[0,\dfrac{\pi}{2}\right]$; (2) $[0,\pi]$; (3) $\left[0,\dfrac{3}{2}\pi\right]$.

解 (1) 在 $\left[0,\dfrac{\pi}{2}\right]$ 上, $f(x)=\sin x\geq 0$

且 $$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=\int_{0}^{\frac{\pi}{2}}\sin x\mathrm{d}x=1$$

故 $\sin x$ **可作为概率密度.即**

$$f(x)=\begin{cases}\sin x, & 0\leq x\leq\dfrac{\pi}{2};\\ 0, & \text{其他}.\end{cases}$$

(2) **在** $[0,\pi]$ **上,在子区间** $\left[\dfrac{\pi}{2},\pi\right]$ **上**, $\sin x<0$, **故** $\sin x$ **不能作为概率密度**.

(3) **在** $\left[0,\dfrac{3}{2}\pi\right]$ **上,在子区间** $\left[\dfrac{\pi}{2},\pi\right]$ **上**, $\sin x<0$, **故** $\sin x$ **不能作为概率密度**.

(2-9) **连续型随机变量** X **的概率密度为**

$$f(x)=\begin{cases}A\cos x, & |x|<\dfrac{\pi}{2};\\ 0, & |x|\geq\dfrac{\pi}{2}.\end{cases}$$

求:(1) **系数** A;

(2) **随机变量** X **落在区间** $\left(0,\dfrac{\pi}{4}\right)$ **内的概率**.

解 (1) **由** $\displaystyle\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}A\cos x\mathrm{d}x=2A\int_{0}^{\frac{\pi}{2}}\cos x\mathrm{d}x$

$$=2A=1,$$

故 $$A=\frac{1}{2}.$$

(2) $P\left\{0<X<\dfrac{\pi}{4}\right\}=\displaystyle\int_{0}^{\frac{\pi}{4}}\frac{1}{2}\cos x\mathrm{d}x=\frac{\sqrt{2}}{4}$.

(2-10) **设连续随机变量** X **的概率密度为**

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

求随机变量 X 落在区间 $(0.4,1.2)$ 内的概率.

解 $P\{0.4<X<1.2\}=\int_{0.4}^{1.2}f(x)\mathrm{d}x$

$$=\int_{0.4}^{1}x\mathrm{d}x+\int_{1}^{1.2}(2-x)\mathrm{d}x$$

$$=\frac{x^2}{2}\Bigg|_{0.4}^{1}+\left(2x-\frac{x^2}{2}\right)\Bigg|_{1}^{1.2}$$

$$=0.42+0.18=0.6.$$

(2-11) 设某种元件的寿命(以小时计)的概率密度为

$$f(x)=\begin{cases} \dfrac{1\ 000}{x^2}, & x\ge 1\ 000;\\ 0, & x<1\ 000.\end{cases}$$

一台设备中装有 3 个这样的元件.求:

(1) 最初 1 500 小时内没有 1 个元件损坏的概率;

(2) 最初 1 500 小时内只有 1 个元件损坏的概率.

解 1 个元件在 1 500 小时内损坏的概率为

$$P\{X<1\ 500\}=\int_{1\ 000}^{1\ 500}\frac{1\ 000}{x^2}\mathrm{d}x=-\frac{1\ 000}{x}\Bigg|_{1\ 000}^{1\ 500}=\frac{1}{3}.$$

设 Y 表示这台设备 3 个元件中使用 1 500 小时内损坏的个数,显然 $Y\sim B\left(3,\dfrac{1}{3}\right)$.

(1) $P\{Y=0\}=C_3^0\cdot\left(\dfrac{1}{3}\right)^0\cdot\left(\dfrac{2}{3}\right)^3=\dfrac{8}{27}$.

(2) $P\{Y=1\}=C_3^1\cdot\left(\dfrac{1}{3}\right)^1\cdot\left(\dfrac{2}{3}\right)^2=\dfrac{4}{9}$.

(2-12) 设随机变量 X 的分布函数为

$$F(x)=\begin{cases} 0, & x\le 0;\\ Ax^2, & 0< x\le 1;\\ 1, & x>1.\end{cases}$$

求:(1) 系数 A;

(2) $P\{0.3\leq X\leq 0.7\}$.

解 (1) 分布函数 $F(x)$总是右连续的,在 $x=1$, $\lim\limits_{x\to 1+0} F(x)=1$,而

$$F(1) = Ax^2\Big|_{x=1} = A,$$

故
$$A=1.$$

(2) $P\{0.3\leq X\leq 0.7\}=F(0.7)-F(0.3)$

$$=0.49-0.09=0.4.$$

(2-13) 连续型随机变量 X 的分布函数为

$$F(x) = A + B\arctan x, \quad -\infty < x < +\infty.$$

求:(1) 常数 A, B;

(2) $P\{-1< X<1\}$.

解 (1) $F(-\infty)=0=A-\dfrac{\pi}{2}B$. ①

$$F(+\infty)=1=A+\frac{\pi}{2}B \qquad ②$$

联立①,②式可求得 $A=\dfrac{1}{2}$, $B=\dfrac{1}{\pi}$.

故
$$F(x)=\frac{1}{2}+\frac{1}{\pi}\arctan x, \quad -\infty<x<+\infty.$$

(2) $P\{-1< X<1\}=P\{-1< X\leq 1\}=F(1)-F(-1)$

$$=\left(\frac{1}{2}+\frac{1}{\pi}\arctan x\right)\Bigg|_{-1}^{1}$$

$$=\frac{1}{4}-\left(-\frac{1}{4}\right)=\frac{1}{2}.$$

(2-14) 设连续型随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\pi\sqrt{1-x^2}}, & |x|<1;\\ 0, & |x|\geq 1.\end{cases}$$

求 X 的分布函数.

解　当 $x < -1$ 时

$$F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u = \int_{-\infty}^{x} 0\mathrm{d}u = 0;$$

当 $-1 \leqslant x < 1$ 时

$$F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u = \int_{-1}^{x} \frac{1}{\pi\sqrt{1-u^2}}\mathrm{d}u$$

$$= \frac{1}{\pi}\arcsin u \Big|_{-1}^{x} = \frac{1}{\pi}\arcsin x + \frac{1}{2};$$

当 $x \geqslant 1$ 时

$$F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u = \int_{-1}^{1} \frac{1}{\pi\sqrt{1-u^2}}\mathrm{d}u + \int_{1}^{x} 0\mathrm{d}u = 1 .$$

故

$$F(x) = \begin{cases} 0, & x < -1; \\ \dfrac{1}{\pi}\arcsin x + \dfrac{1}{2}, & -1 \leqslant x < 1; \\ 1, & x \geqslant 1 . \end{cases}$$

(2－15)　**测量某目标的距离时发生的随机误差** X **(单位:m)具有概率密度**

$$f(x) = \frac{1}{40\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{3\,200}} .$$

求在 3 次测量中至少有 1 次误差的绝对值不超过 30 m 的概率 .

解　由题设, $X \sim N(0, 40^2)$, 故在 1 次测量中误差的绝对值不超过 30 m 的概率为

$$P\{|X| \leqslant 30\} = P\{-30 \leqslant X \leqslant 30\} = \Phi\left(\frac{30}{40}\right) - \Phi\left(-\frac{30}{40}\right)$$

$$= \Phi(0.75) - \Phi(-0.75) = 2\Phi(0.75) - 1$$

$$= 2 \times 0.773\,4 - 1 = 0.546\,8 .$$

由此知,3 次测量中至少有 1 次误差的绝对值不超过 30 m 的概率为

$$p = 1 - (1 - 0.546\,8)^2 = 0.906\,9 .$$

(2－16)　**公共汽车车门高度是按男子与车门顶碰头的机会**

在 0.01 以下来设计的.设男子身高服从 $N(170, 6^2)$(单位: cm),试确定车门的高度.

解　设车门设计高度为 l cm,根据设计要求,设男子身高为 X cm,应有

$$P\{X > l\} < 0.01,$$

而 $$P\{X > l\} = 1 - P\{X \leq l\} = 1 - \Phi\left(\frac{l - 170}{6}\right) < 0.01,$$

即有 $$\Phi\left(\frac{l - 170}{6}\right) > 0.99,$$

查附表 1 得 $$\frac{l - 170}{6} > 2.33,$$

即车门高度应大于 183.98≈184 cm.

(2－17)　设离散随机变量 X 的分布列为

X	-2	-1	0	1	2	3
p	0.10	0.20	0.25	0.20	0.15	0.10

求:(1) $Y_1 = -2X$ 的分布列;

(2) $Y_2 = X^2$ 的分布列.

解　(1) Y_1 的分布列为

Y_1	4	2	0	-2	-4	-6
p	0.10	0.20	0.25	0.20	0.15	0.10

(2) Y_2 的可能值为 0,1,4,9.如

$$\begin{aligned} P\{X^2 = 4\} &= P\{(X = 2) \cup (X = -2)\} \\ &= P\{X = 2\} + P\{X = -2\} \\ &= 0.15 + 0.1 = 0.25. \end{aligned}$$

得 Y_2 的分布列为

Y_2	0	1	4	9
p	0.25	0.40	0.25	0.10

(2－18)　(1) 设连续随机变量 X 的概率密度为 $f_X(x)$,求 Y

$=X^3$ 的概率密度 .

(2) 设 X 服从指数分布 $E(\lambda)$, 求 $Y=X^3$ 的概率密度 .

解 (1) $y=x^3$ 为单调增函数,反函数为 $x=y^{\frac{1}{3}}$,故

$$f_Y(y)=f_X(y^{\frac{1}{3}})\frac{1}{3}y^{-\frac{2}{3}}=\frac{1}{3}f_X(\sqrt[3]{y})y^{-\frac{2}{3}}\quad(y\neq 0).$$

(2) $f_X(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & x\leq 0.\end{cases}$ 利用(1)的结果,有

$$f_Y(y)=\begin{cases}\frac{1}{3}\lambda e^{-\lambda\sqrt[3]{y}}y^{-\frac{2}{3}}, & y>0;\\ 0, & y\leq 0.\end{cases}$$

(2-19) 设 X 服从标准正态分布 $N(0,1)$, 求 $Y=2X^2+1$ 的概率密度 .

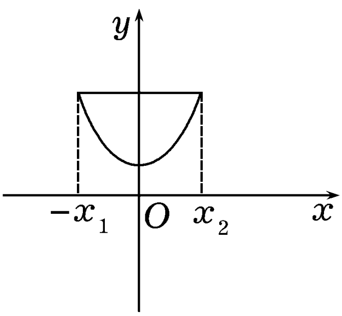


图 2-1

解 $y=2x^2+1$, $x^2=\frac{y-1}{2}$.如图 2-1,当 $y\geq 1$ 时, Y 的分布函数

$$F_Y(y)=P(Y\leq y)=P\left\{X^2\leq\frac{y-1}{2}\right\}=P\{-x_1\leq X\leq x_1\},$$

其中 $x_1=\sqrt{\frac{y-1}{2}}$.从而

$$F_Y(y)=\int_{-x_1}^{x_1}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\int_{-\sqrt{\frac{y-1}{2}}}^{\sqrt{\frac{y-1}{2}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$$=2\int_0^{\sqrt{\frac{y-1}{2}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx.$$

对变上限求导,得

$$f_Y(y)=\frac{1}{2\sqrt{\pi(y-1)}}e^{-\frac{y-1}{4}}.$$

当 $y<1$ 时, $F_Y(y)=P\{Y\leq y\}=0$,从而 $f_Y(y)=0$,

故
$$f_Y(y)=\begin{cases}\dfrac{1}{2\sqrt{\pi(y-1)}}\mathrm{e}^{-\frac{y-1}{4}}, & y\geq 1;\\ 0, & y<1.\end{cases}$$

(2－20) 设电压 $V=A\sin\Theta$，其中 A 是一个正常数，相角 Θ 是一个随机变量，服从均匀分布 $U\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$，求电压 V 的概率密度．

解 Θ 的概率密度

$$f_\Theta(\theta)=\begin{cases}\dfrac{1}{\pi}, & -\dfrac{\pi}{2}<\theta<\dfrac{\pi}{2};\\ 0, & \text{其他}.\end{cases}$$

函数 $V=A\sin\theta$ 在 $\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$ 上单调增加，其反函数 $\theta=h(V)=\arcsin\dfrac{V}{A}$ 单值．

当 $|v|\geq A$ 时，V 的概率密度 $f_V(v)=0$．

当 $|v|<A$（即 $-A<v<A$）时，

$$h'(v)=\frac{1}{\sqrt{1-\left(\dfrac{v}{A}\right)^2}}\cdot\frac{1}{A}=\frac{1}{\sqrt{A^2-v^2}}$$

$$f_V(v)=\frac{1}{\pi}\left|h'(v)\right|=\frac{1}{\pi\sqrt{A^2-v^2}}.$$

故
$$f_V(v)=\begin{cases}\dfrac{1}{\pi\sqrt{A^2-v^2}}, & -A<v<A;\\ 0, & \text{其他}.\end{cases}$$

（补 2－1） 某种型号的电子管的寿命 X（以小时计）具有以下的概率密度：

$$f(x)=\begin{cases}\dfrac{1\ 000}{x^2}, & x>1\ 000;\\ 0, & \text{其他}.\end{cases}$$

现有一大批此种管子（设各电子管损坏与否相互独立），任取 5 只，问其中至少有 2 只寿命大于 1 500 小时的概率是多少？

解　$P\{X\geq 1\,500\}=\int_{1\,500}^{+\infty}\frac{1\,000}{x^2}\mathrm{d}x=\frac{2}{3}$，记 Y 为任取 5 只，其寿命大于 1 500 小时的电子管数．

则

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

$$P\{Y\geq 2\}=1-P(0)-P(1)$$

$$=1-\left(\frac{1}{3}\right)^5-5\left(\frac{1}{3}\right)^4\times\frac{2}{3}$$

$$=1-\frac{1+10}{243}=\frac{232}{243}=0.954\,7.$$

（补 2－2）　工厂生产的电子管的寿命 X（以小时计）服从参数为 $\mu=160$，σ 的正态分布，若要求 $P\{120<X\leq 200\}\geq 0.80$，允许 σ 最大为多少？

解　由已知，$X\sim N(160,\sigma^2)$，则

$$\frac{X-160}{\sigma}\sim N(0,1).$$

$$P\{120<X\leq 200\}=P\left\{-\frac{40}{\sigma}<\frac{X-160}{\sigma}\leq\frac{40}{\sigma}\right\}$$

$$=\Phi\left(\frac{40}{\sigma}\right)-\Phi\left(-\frac{40}{\sigma}\right)=2\Phi\left(\frac{40}{\sigma}\right)-1.$$

若要求 $P\{120<X\leq 200\}\geq 0.80$，则要求

$$2\Phi\left(\frac{40}{\sigma}\right)-1\geq 0.80,\quad \Phi\left(\frac{40}{\sigma}\right)\geq 0.9.$$

查表得 $\frac{40}{\sigma}\geq 1.29$，所以 $\sigma\leq 31.25$，即 σ 最大为 31.25．

试 题 选 解

[2－1]　填空题

1．设随机变量 X 服从参数为 λ 的泊松分布，且 $P(X=1)=P(X=2)$，则 $P(X\geq 3)=$ ________．

答　$1-5e^{-2}$.

因为 $X\sim P(\lambda)$, $P(X=i)=\dfrac{\lambda^i}{i!}e^{-\lambda}$. 由 $P(X=1)=P(X=2)$ $(\lambda>0)$ 得 $\lambda=2$. 所以 $P(X\geq 3)=1-P(X<3)=1-P(X=0)-P(X=1)-P(X=2)=1-5e^{-2}$.

2. 设 $X\sim N(10, 0.02^2)$. 已知 $\Phi(2.5)=0.9938$, 则概率 $P(9.95<X<10.05)=$ ________ .

答　0.9876 .

因为

$$
\begin{aligned}
P(9.95<X<10.05)&=\Phi\left(\frac{10.05-10}{0.02}\right)-\Phi\left(\frac{9.95-10}{0.02}\right)\\
&=2\Phi(2.5)-1=2\times 0.9938-1\\
&=1.9876-1=0.9876.
\end{aligned}
$$

3. 设 X 的概率分布为

$$
P(X=K)=\frac{a}{k(k+1)}\quad (k=1,2,\cdots).
$$

则常数 $a=$ ________ .

答　1 .

由分布列性质

$$
\sum_{k=1}^{+\infty}p_k=\sum_{k=1}^{+\infty}P\{X=k\}=1,
$$

得 $1=\displaystyle\sum_{k=1}^{+\infty}\frac{a}{k(k+1)}=a\sum_{k=1}^{+\infty}\frac{1}{k(k+1)}=a$, 故 $a=1$.

4. 设 $X\sim N(10,9)$. 已知 $\Phi(1)=0.8413$, 则 $P(7<X\leq 13)$ $=$ ________ .

答　0.6826 .

因为

$$
\begin{aligned}
P(7<X\leq 13)&=\Phi\left(\frac{13-10}{3}\right)-\Phi\left(\frac{7-10}{3}\right)\\
&=\Phi(1)-\Phi(-1)=2\Phi(1)-1\\
&=2\times 0.8413-1\\
&=1.6826-1=0.6826.
\end{aligned}
$$

5．设在一次试验中 A 出现的概率为 $\frac{1}{3}$，则在 3 次重复独立试验中 A 至少出现 1 次的概率为________．

答　$\frac{19}{27}$．

设 X 为 3 次重复独立试验中 A 出现的次数，依题意 $X \sim B(3, p)$，$p=\frac{1}{3}$，即 $X \sim B\left(3, \frac{1}{3}\right)$．

故所求概率为 $P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - \left(\frac{2}{3}\right)^3 = \frac{19}{27}$．

6．已知随机变量 X 的分布列为

X	1	2	3	4	5
p	$2a$	0.1	0.3	a	0.3

则常数 $a=$________．

答　0.1．

依分布列性质 $2a + 0.1 + 0.3 + a + 0.3 = 3a + 0.7 = 1$，故 $a = 0.1$．

7．设随机变量 $X \sim N(0,1)$，$\Phi(x)$ 为其分布函数，则 $\Phi(x) + \Phi(-x) =$________．

答　1．

因为 $\Phi(x) + \Phi(-x) = \Phi(x) + 1 - \Phi(x) = 1$．

8．已知连续型随机变量 X 的分布函数为

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

设 X 的概率密度为 $f(x)$，则当 $x < 0$ 时，$f(x) =$________．

答　$\frac{1}{3}e^x$．

由于在 $f(x)$ 的连续点处有 $\frac{\mathrm{d}}{\mathrm{d}x}F(x)=f(x)$，所以当 $x<0$ 时，$f(x)=\left(\frac{1}{3}\mathrm{e}^{x}\right)'=\frac{1}{3}\mathrm{e}^{x}$.

[2-2] 单选题

1．已知随机变量 X 的概率密度为 $f_X(x)$，令 $Y=-2X$，则 Y 的概率密度 $f_Y(y)$ 为 ( )

① $2f_X(-2y)$ ② $f_X\left(-\frac{y}{2}\right)$

③ $-\frac{1}{2}f_X\left(-\frac{y}{2}\right)$ ④ $\frac{1}{2}f_X\left(-\frac{y}{2}\right)$

答 选择④ .

因为 $y=-2x$ 为单调函数，它的反函数为 $x=-\frac{y}{2}$，$x'_y=-\frac{1}{2}$，则由定理知 $Y=-2X$ 的概率密度为

$$f_Y(y)=f_X\left(-\frac{y}{2}\right)\left|-\frac{1}{2}\right|=\frac{1}{2}f_X\left(-\frac{y}{2}\right).$$

2．如果函数 $f(x)=\begin{cases}x, & a\leq x\leq b;\\ 0, & x<a \text{ 或 } x>b\end{cases}$ 是某连续型随机变量 X 的概率密度，则区间 $[a,b]$ 可以是 ( )

① $[0,1]$ ② $[0,2]$ ③ $[0, 2]$ ④ $[1,2]$

答 选择③ .

由性质 $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$ 得 $\int_a^b x\mathrm{d}x=1$，即 $b^2-a^2=2$，由此可证明①，②，④均不成立 .

3．下列各函数是随机变量分布函数的为 ( )

① $F_1(x)=\frac{1}{1+x^2}$， $-\infty<x<+\infty$

② $F_2(x)=\begin{cases}0, & x\leq 0,\\ \frac{x}{1+x}, & x>0\end{cases}$

③ $F_3(x)=\mathrm{e}^{-x}$， $-\infty<x<+\infty$

④ $F_4(x)=\frac{3}{4}+\frac{1}{2\pi}\arctan x,\quad -\infty<x<+\infty$

答　选择② .

由随机变量分布函数的 3 条基本性质, 如 $F(-\infty)=0$, $F(+\infty)=1$可证明①,③,④均不是分布函数 .

[2 - 3]　是非题

1 . 分布函数都是连续函数 .　　　　　　　　　　　　　　　　(　　)

答　不正确 .

因为对连续型随机变量, $F(x)$是连续函数;对离散型随机变量, $F(x)$右连续 .

2 . 设 X 的密度函数为 $f(x)$, 则 $2X$ 的密度函数为$\frac{1}{2} f \quad \frac{y}{2}$.

(　　)

答　正确 .

因为根据连续型随机变量的函数概率密度定理,可知 $Y=2X$ 的密度函数为

$$f_Y(y) = f_X(h(y))\mid h'(y)\mid = \frac{1}{2} f \quad \frac{x}{2} \quad ,$$

其中
$$h(y)=\frac{y}{2},\quad h'(y)=\frac{1}{2} .$$

3 . 若对随机变量 X 有 $P\{X=k\}=\frac{1}{k+1}$, $k=1,2,3$, 则它是 X 的概率分布 .　　　　　　　　　　　　　　　　(　　)

答　不正确 .

因为 $P(X=1)+P(X=2)+P(X=3)\neq 1$.

4 . 某型号电阻次品率为$\frac{1}{3}$,现从中任取 4 个,则这 4 个中仅有 1 个次品的概率为$\frac{32}{81}$.

答　正确 .

设 X 为 4 个中的次品数 .由于 $X\sim B\quad 4,\frac{1}{3}$,故

$$P\{X=1\}=C_4^1\times\frac{1}{3}\times\left(\frac{2}{3}\right)^3=\frac{32}{81}.$$

5. 连续型随机变量的分布密度函数 $f(x)$ 一定是 x 的连续函数.

答　不正确.

例如均匀分布.

6. 设 X 服从参数为 $\lambda=2$ 的泊松分布,则

$$P(X<3)=4e^{-2}.$$

答　不正确.

因为 $P(X<3)=P(X=0)+P(X=1)+P(X=2)$

$$=e^{-2}+2e^{-2}+\frac{2^2}{2!}e^{-2}=5e^{-2}.$$

[2－4]　设顾客在某银行的窗口等待服务的时间(以分钟计)服从指数分布,其密度为

$$f_X(x)=\begin{cases}\frac{1}{5}e^{-\frac{x}{5}}, & x>0;\\ 0, & \text{其他}.\end{cases}$$

某顾客若等待时间超过 10 分钟他就离开,一个月他去银行 5 次,以 Y 表示一个月内他未等到服务而离开的次数.写出 Y 的分布律并求 $P(Y\geq1)$.

解　设 A 表示"某顾客到银行未等到服务",则

$$P(A)=P\{X>10\}=1-P\{X\leq10\}=1-\int_{-\infty}^{10}f(x)\mathrm{d}x$$

$$=1-\int_0^{10}\frac{1}{5}e^{-\frac{x}{5}}\mathrm{d}x=1-1+e^{-2}=e^{-2}.$$

一个月顾客去银行 5 次,就是做 5 次独立试验,每次去银行而未得到服务的概率为 e^{-2},因此 $Y\sim B(5,e^{-2})$,Y 的分布律为

$$P\{Y=k\}=C_5^k e^{-2k}(1-e^{-2})^{5-k},\quad k=0,1,2,3,4,5.$$

$$P\{Y\geq1\}=1-P\{Y=0\}=1-(1-e^{-2})^5.$$

分析　本题的关键要观察出 $Y\sim B(5,e^{-2})$.因此,首先就得

求出 $P(A)$.

[2－5]　随机变量 X 服从 $(0,1)$ 上的均匀分布，求 $Y=-2\ln X$ 的密度函数.

解法一　由已知得 X 的概率密度为

$$f_X(x)=\begin{cases}1, & x\in(0,1);\\ 0, & \text{其他}.\end{cases}$$

因为 $y=-2\ln x$ 处处可导，且当 $x\in(0,1)$ 时，$y'=-\dfrac{2}{x}<0$. $y=-2\ln x$ 的反函数为 $x=e^{-\frac{y}{2}}$，$x'=-\dfrac{1}{2}e^{-\frac{y}{2}}$.另外，当 X 在 $(0,1)$ 中取值时，Y 在 $(0,+\infty)$ 中取值，所以 $Y=-2\ln X$ 的概率密度为

$$f_Y(y)=\begin{cases}\dfrac{1}{2}e^{-\frac{y}{2}}, & y>0;\\ 0, & y\leq 0.\end{cases}$$

解法二　Y 的分布函数

$$F(y)=P\{-2\ln X\leq Y\}=P\{X\geq e^{-\frac{y}{2}}\}$$

当 $y\leq 0$ 时，　$e^{-\frac{y}{2}}\geq 1$，　$F(y)=0$；

而当 $y>0$ 时，　$e^{-\frac{y}{2}}\leq 1$，　$F(y)=1-e^{-\frac{y}{2}}$.

从而

$$f_Y(y)=\begin{cases}\dfrac{1}{2}e^{-\frac{y}{2}}, & y>0;\\ 0, & y\leq 0.\end{cases}$$

[2－6]　设 $X\sim N(0,1)$，求 $Y=X^2$ 的分布密度函数.

解　因为 $X\sim N(0,1)$，所以

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

设 Y 的分布函数为 $F_Y(y)$，则

当 $y\leq 0$ 时，

$$f_Y(y)=P(Y\leq y)=P(X^2\leq y)=0,$$

故
$$f_Y(y) = F'_Y(y) = 0 .$$

当 $y > 0$ 时,

$$f_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(|X| \leq \sqrt{y})$$
$$= P(-\sqrt{y} \leq X \leq \sqrt{y}) .$$
$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx .$$

故
$$f_Y(y) = F'_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot y^{-\frac{1}{2}} .$$

从而
$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot y^{-\frac{1}{2}}, & y > 0; \\ 0, & y \leq 0 . \end{cases}$$

第三章　随 机 向 量

内 容 提 要

一、知识脉络

- 二维随机向量
  - 离散型
    - 定义
    - 联合分布列 — 边缘分布列
  - 连续型
    - 定义
    - 联合概率密度 — 边缘概率密度
    - 常见二维随机变量的分布 — 二维均匀分布、二维正态分布
  - 两个随机变量的独立性
    - 定义
    - 二维随机变量之和的概率分布
- n 维随机向量
  - 定义
  - n 种常用的分布
    - χ^2 分布
    - t 分布
    - F 分布

二、知识要点

(一) 二维随机向量及其分布函数

1．二维随机向量的定义

设 E 是一随机试验，Ω 是其样本空间，$X=X(\omega)$，$Y=Y(\omega)$ 是定义在 Ω 上的随机变量，则称 (X,Y) 是二维随机向量或二维随机变量．

2．分布函数

设 (X,Y) 是一个二维随机向量，对任意实数 x,y，令 $F(x,y)=P\{X\leq x,Y\leq y\}$，$-\infty<x<+\infty$，$-\infty<y<+\infty$．称二元函数 $F(x,y)$ 为 (X,Y) 的联合分布函数，简称为 (X,Y) 的分布函数．

3．分布函数 $F(x,y)$ 的性质

(1) $F(x,y)$ 是变量 x 和 y 的不减函数；

(2) 对一切 x,y，$0\leq F(x,y)\leq 1$，且

$$F(-\infty,y)=\lim_{x\to-\infty}F(x,y)=0,$$

$$F(x,-\infty)=\lim_{y\to-\infty}F(x,y)=0,$$

$$F(+\infty,+\infty)=\lim_{\substack{x\to+\infty\\y\to+\infty}}F(x,y)=1;$$

(3) $F(x,y)$ 关于 x 右连续，关于 y 也右连续；

(4) 对任意 $x_1<x_2$，$y_1<y_2$，都有

$$\begin{aligned}&P\{x_1<X\leq x_2,y_1<Y\leq y_2\}\\&=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\geq 0.\end{aligned}$$

(二) 二维离散型随机向量的分布列

1．联合分布列

设 (X,Y) 是二维离散型随机向量，其所有可能的取值为 (x_i,y_i)，$i,j=1,2,\cdots$，则称

$$P\{X=x_i,Y=y_i\}=P_{ij},\quad i,j=1,2,\cdots.$$

为二维离散型随机向量(X, Y)的联合分布列，简称为(X, Y)的分布列.

2. 分布列 P_{ij} 的性质

(1) $P_{ij} \geq 0$;

(2) $\sum_i \sum_j P_{ij} = 1$.

若已知(X, Y)的分布列为

$$P\{X = x_i, Y = y_j\} = P_{ij}, \ i, j = 1, 2, \cdots,$$

则(X, Y)的分布函数为

$$F(x, y) = \sum_{y_j \leq y} \sum_{x_i \leq x} P\{X = x_i, Y = y_j\}.$$

(三) 二维连续型随机向量的联合概率密度

1. 定义

设(X, Y)是二维随机向量，$F(x, y)$是其分布函数，若存在二元函数 $f(x, y) \geq 0$，使得对于任意实数 x, y 都有

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

则称 $f(x, y)$为(X, Y)的联合概率密度.

2. 联合概率密度的性质

(1) $f(x, y) \geq 0$;

(2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \mathrm{d}x \mathrm{d}y = 1$;

(3) 若 $f(x, y)$在(x, y)处连续，则 $f(x, y) = \dfrac{\partial^2 F(x, y)}{\partial x \partial y}$;

(4) 设 G 是 xoy 平面上任一区域，则

$$P\{(x, y) \in G\} = \iint_G f(x, y) \mathrm{d}x \mathrm{d}y.$$

(四) 边缘分布

1. 定义

对于二维随机向量(X, Y)，令

$$F_X(x)=P\{X\leq x\}=F(x,+\infty),\ x\text{是任意实数};$$
$$F_Y(y)=P\{Y\leq y\}=F(+\infty,y),\ y\text{是任意实数}.$$
则称 $F_X(x)$, $F_Y(y)$分别为(X,Y)关于 X 和 Y 的边缘分布函数.

2. 离散型情形

已知(X,Y)的联合分布列为 $P\{X=x_i,Y=y_j\}=P_{ij}$, $\quad i,j=1,2,\cdots$,记

$$P\{X=x_i\}=\sum_{j=1}^{+\infty}P\{X=x_i,Y=y_j\}=\sum_{j=1}^{+\infty}P_{ij}=P_{i\cdot},i=1,2,\cdots,$$
$$P\{Y=y_j\}=\sum_{i=1}^{+\infty}P\{X=x_i,Y=y_j\}=\sum_{i=1}^{+\infty}P_{ij}=p_{\cdot j},j=1,2,\cdots,$$

则称 $p_{i\cdot}$, $p_{\cdot j}(i,j=1,2,\cdots)$分别为$(X,Y)$关于 X 和 Y 的边缘分布列.

3. 连续型情形

已知(X,Y)只有联合概率密度 $f(x,y)$,记

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y,$$
$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x.$$

称 $f_X(x)$, $f_Y(y)$分别为(X,Y)关于 X 和 Y 的边缘概率密度.

(五) 随机变量的独立性

定义 设 $F(x,y)$和 $F_X(x)$, $F_Y(y)$分别是(X,Y)的联合分布函数和边缘分布函数,若对任意实数 x,y,都有

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

则称随机变量 X 与 Y 是相互独立的.

若(X,Y)是离散型随机向量,则

$$X,Y\text{ 相互独立}\quad P\{X=x_i,Y=y_j\}=P\{X=x_i\}\cdot P\{Y=y_j\}.$$

若(X,Y)是连续型随机向量,则

$$X,Y\text{ 相互独立}\quad f(x,y)=f_X(x)\cdot f_Y(y).$$

(六) 两个随机变量和的分布

1 . 离散型情形

设离散型随机变量 X **与** Y **相互独立,** X, Y **的可能取值均为非负整数** $0,1,2,\cdots$**,则** $Z = X + Y$ **的可能取值也是非负整数** $0,1,2,\cdots$ **.** Z **的分布律为**

$$
\begin{aligned}
P\{Z = k\} &= P\{X + Y = k\} \\
&= \sum_{i=0}^{K} P\{X = i\} \cdot P\{Y = k - i\}, \quad k = 0,1,2,\cdots .
\end{aligned}
$$

2 . 连续型情形

已知 (X, Y) **具有联合概率密度** $f(x, y)$**,则** $Z = X + Y$ **有概率密度**

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

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

特别是,若 X, Y **相互独立,则有**

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

或
$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y)\mathrm{d}y .$$

(七) n **维随机向量**

(1) **设** $(X_1, X_2, \cdots, X_n)$ **为** n **维相互独立的连续型随机变量,** $f(x_1, x_2, \cdots, x_n)$ **为其连续的联合概率密度,则**

$$f(x_1, x_2, \cdots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2)\cdots f_{X_n}(x_n) .$$

(2) χ^2 **分布　设** $X_1, X_2, \cdots, X_n$ **相互独立,且都服从标准正态分布** $N(0,1)$**,定义随机变量**

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2,$$

称随机变量 χ^2 **服从自由度为** n **的** χ^2 **分布,记为** $\chi^2 \sim \chi^2(n)$.

(3) t **分布　设** $X \sim N(0,1)$, $Y \sim \chi^2(n)$**,并且** X **与** Y **相互独立,则称随机变量**

$$t = \frac{X}{Y/\ n}$$

服从自由度为 n **的** t **分布,记为** $t \sim t(n)$.

(4) F **分布**　**设** $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$,**且** X **与** Y **相互独立,则称随机变量**

$$F = \frac{X/\ n_1}{Y/\ n_2}$$

服从自由度为(n_1, n_2)**的** F **分布,记为** $F \sim F(n_1, n_2)$.

要 点 范 例

一、二维离散型随机向量分布列、边缘分布列及分布函数的计算

例 3 1　**设随机变量** X **在** 1,2,3,4 **四个整数中等可能地取值,另一个随机变量** Y **在** $1 \sim X$ **中等可能地取值,试求**(X, Y)**的分布列** .

分析　**先将两个随机变量所有可能的取值列出,再考虑取每一对数值时的概率** .

解　X **的可能取值为** 1,2,3,4, Y **的取值不能超过** X **的取值,即在**$\{X = i\}$**发生的条件下**, Y **只可能取小于或等于** i **的一个整数,故**

$$P\{X = i, Y = j\} = P\{X = i\} \cdot P\{Y = j \mid X = i\}$$

$$= \frac{1}{4} \cdot \frac{1}{i} = \frac{1}{4i}, \quad i = 1,2,3,4 \quad (j \leq i) .$$

当 $j > i$ **时**,$\{X = i, Y = j\}$**为不可能事件** .**因此有**

$$P\{X = i, Y = j\} = 0 \quad (j > i),$$

于是得(X, Y)**的分布列为**

X \ Y	1	2	3	4
1	$\frac{1}{4}$	0	0	0
2	$\frac{1}{8}$	$\frac{1}{8}$	0	0
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	0
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

例 3.2 射手进行射击,击中目标的概率为 $p(0 < p < 1)$,射击进行到击中目标两次为止.设 X 表示第一次击中目标所进行的射击次数,Y 表示总共进行的射击次数.

(1) 求(X, Y)的联合分布列;

(2) 求(X, Y)的边缘分布列.

解 (1) 随机变量 X 的可能取值为 $1, 2, 3, \cdots$,随机变量 Y 的可能取值为 $2, 3, 4, \cdots$,(X, Y)的联合分布列为 $P\{X = m, Y = n\} = p^2(1-p)^{n-2}(m = 1, 2, \cdots; n = m+1, m+2, \cdots)$.

(2) 先求关于 X 的边缘分布列

$$\begin{aligned} P\{X = m\} &= \sum_{n=m+1}^{+\infty} P\{X = m, Y = n\} \\ &= \sum_{n=m+1}^{+\infty} p^2(1-p)^{n-2} = \frac{p^2}{(1-p)^2}\sum_{n=m+1}^{+\infty}(1-p)^n \\ &= \frac{p^2}{(1-p)^2}\cdot\frac{(1-p)^{m+1}}{1-(1-p)} = p(1-p)^{m-1}, \\ & \qquad m = 1, 2, 3, \cdots. \end{aligned}$$

再求关于 Y 的边缘分布列

$$\begin{aligned} P\{Y = n\} &= \sum_{m=1}^{n-1} P\{X = m, Y = n\} = \sum_{m=1}^{n-1} p^2(1-p)^{n-2} \\ &= (n-1)p^2(1-p)^{n-2}, \quad n = 2, 3, 4, \cdots. \end{aligned}$$

例 3.3 设随机向量(X, Y)的联合分布列为

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

(1) 求(X, Y)的联合分布函数 $F(x, y)$;

(2) 求 $P\{X\leq 2.5, Y\leq 1.5\}$, $P\{X<2, Y<2\}$.

解 (1) 如图 3-1 所示,随机点的连线把 $F(x, y)$的定义域(xOy 平面)分成 9 个区域,显然有

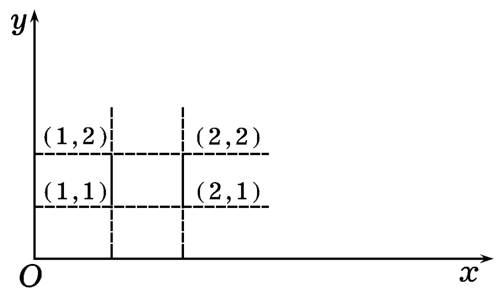


图 3-1

当 $x<1$ 或 $y<1$ 时, $F(x, y)=P\{X\leq x, Y\leq y\}=P\{\varnothing\}=0$,这相当于考虑了其中 5 个区域的概率,所以还需考虑余下 4 个区域内的概率;

当 $1\leq x<2, 1\leq y<2$ 时, $F(x, y)=p_{11}=0$;

当 $1\leq x<2, y\geq 2$ 时, $F(x, y)=p_{11}+p_{12}=\dfrac{1}{2}$;

当 $x\geq 2, 1\leq y<2$ 时, $F(x, y)=p_{11}+p_{21}=\dfrac{1}{4}$;

当 $x\geq 2, y\geq 2$ 时, $F(x, y)=p_{11}+p_{12}+p_{21}+p_{22}=1$.

综合上述情况,得随机向量(X, Y)的分布函数为

$$F(x, y)=\begin{cases} 0, & x<1 \text{ 或 } y<1 \text{ 或 } 1\leq x<2 \text{ 且 } 1\leq y<2; \\ \dfrac{1}{2}, & 1\leq x<2, y\geq 2; \\ \dfrac{1}{4}, & x\geq 2, 1\leq y<2; \\ 1, & x\geq 2, y\geq 2. \end{cases}$$

(2) $P\{X\leq 2.5, Y\leq 1.5\} = F\{2.5, 1.5\} = \dfrac{1}{4}$,

$P\{X<2, Y<2\} = F(2,2) - P\{X=2, Y=2\} = 1 - \dfrac{1}{4} = \dfrac{3}{4}$.

注　求二维离散型随机向量落入某一区域的概率时,要特别注意不等号和严格不等号的差别.即 $P\{X\leq a, Y\leq b\}$**与** $P\{X<a, Y<b\}$**表示的概率一般是不相同的.而连续型随机变量无这种差别.**

二、二维连续型随机向量边缘分布密度及分布函数的求法

例 3.4　设二维随机向量(X, Y)**的密度函数为:**

$$f(x,y)=\begin{cases} x^2+\dfrac{1}{3}xy, & 0\leq x\leq 1, 0\leq y\leq 2; \\ 0, & \text{其他}. \end{cases}$$

(1) **求关于** X **和关于** Y **的边缘密度函数;**

(2) **求** $P\{X+Y\geq 1\}$.

解　(1) **求关于** X **的边缘密度函数** $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$,

当 $x<0$ **或** $x>1$ **时**, $f_X(x)=0$;

当 $0\leq x\leq 1$ **时**, $f_X(x)=\int_0^2 \left(x^2+\dfrac{1}{3}xy\right)\mathrm{d}y = 2x^2+\dfrac{2}{3}x$, **所以**

$$f_X(x)=\begin{cases} 2x^2+\dfrac{2}{3}x, & 0\leq x\leq 1; \\ 0, & \text{其他}. \end{cases}$$

求关于 Y **的边缘密度函数** $f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x$.

当 $y<0$ **或** $y>2$ **时**, $f_Y(y)=0$;

当 $0\leq y\leq 2$ **时**, $f_Y(y)=\int_0^1 \left(x^2+\dfrac{1}{3}xy\right)\mathrm{d}x = \dfrac{1}{3}+\dfrac{1}{6}y$, **所以**

$$f_Y(y)=\begin{cases}\dfrac{1}{3}+\dfrac{1}{6}y, & 0\leq y\leq 2;\\ 0, & 其他 .\end{cases}$$

(2) **如图** 3-2 **所示**,

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

$$=\int_0^1 \mathrm{d}x\int_{1-x}^2\left(x^2+\frac{1}{3}xy\right)\mathrm{d}y=\int_0^1\left(x^2y+\frac{1}{6}xy^2\right)\Bigg|_{1-x}^2\mathrm{d}x$$

$$=\left(\frac{5}{24}x^4+\frac{4}{9}x^3+\frac{1}{4}x^2\right)\Bigg|_0^1=\frac{65}{72}.$$

注　**当**(X,Y)**是连续型随机向量时**,**由于** $f(x,y)$**一般是分段函数**,**可按下列方法求边缘密度函数**:**设** $f(x,y)>0$ **的区域为** G,**为求** $f_X(x)$,**可寻求直线** $X=x$ **与** G **截线片段** $D_x=\{y|(x,y)\in G\}$,**那么有** $f_X(x)=\int_{D_x}f(x,y)\mathrm{d}y$.**若直线** $X=x$ **与** G **不相交**,**则** $f_X(x)=0$.**求** $f_Y(y)$**方法类似**.

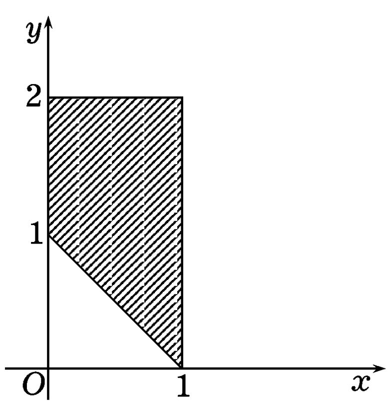


图 3-2

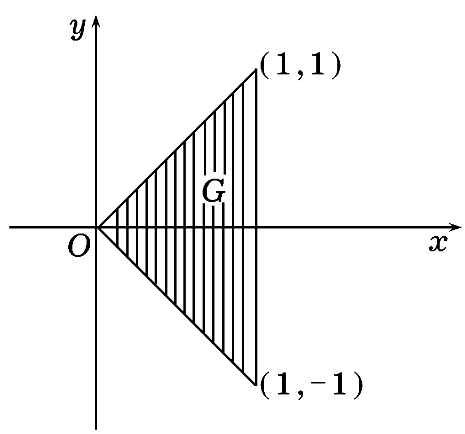


图 3-3

例 3.5　**设随机向量**(X,Y)**在由** $y=x$, $y=-x$,**及** $x=1$ **所围成的三角形区域** G **上服从均匀分布**,**求** X **与** Y **的边缘密度函数**.

解　**由于三角形区域** G **的面积** $A=\dfrac{1}{2}\cdot 2\cdot 1=1$,**所以**$(X,Y)$**的密度函数为**

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

G 可用不等式组表示成: $-x<y\leq x, 0\leq x\leq 1$. 如图 3-3 所示. 为求边缘密度函数 $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$, 应将定义域划分为 3 段: $(-\infty,0), [0,1], (1,+\infty)$.

当 $x<0$ 或 $x>1$ 时, $f_X(x)=0$;

当 $0\leq x\leq 1$ 时, Y 仅在区间 $[-x, x]$ 上 $f(x,y)$ 才有非零值, 故

$$f_X(x)=\int_{-x}^{x}\mathrm{d}x = 2x,$$

于是得到关于 X 的边缘密度函数为

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

为求 $f_Y(y)$, 将 G 分成下列两个互不相交的区域.

$$G_1:\begin{cases}-y\leq x\leq 1,\\ -1\leq y\leq 0;\end{cases}\qquad G_2:\begin{cases}y\leq x\leq 1,\\ 0\leq y\leq 1.\end{cases}$$

$f(x,y)$ 仅在 G_1 和 G_2 上有非零值, 故应将 y 的定义域分为 4 段: $(-\infty,-1), [-1,0), [0,1], (1,+\infty)$.

当 $y<-1$ 或 $y>1$ 时, $f_Y(y)=0$;

当 $-1\leq y<0$ 时, $f_Y(y)=\int_{-y}^{1} 1\mathrm{d}x = 1+y$;

当 $0\leq y\leq 1$ 时, $f_Y(y)=\int_{y}^{1} 1\mathrm{d}x = 1-y$.

综上所述, 得到关于 Y 的边缘密度函数为

$$f_Y(y)=\begin{cases}1+y, & -1\leq y<0;\\ 1-y, & 0\leq y\leq 1;\\ 0, & \text{其他}.\end{cases}$$

例 3.6 设 (X, Y) 的联合概率密度为

$$f(x,y)=\begin{cases}\frac{1}{2}\sin(x+y), & 0\leq x<\frac{\pi}{2}, 0\leq y<\frac{\pi}{2};\\ 0, & \text{其他}.\end{cases}$$

求(X, Y)的联合分布函数.

解 由于联合密度函数 $f(x, y)$是分区域定义的,因此应根据不同区域来计算分布函数 $F(x, y)$. $f(x, y)$仅在 D 内取值不等于零,如图 3-4 所示.

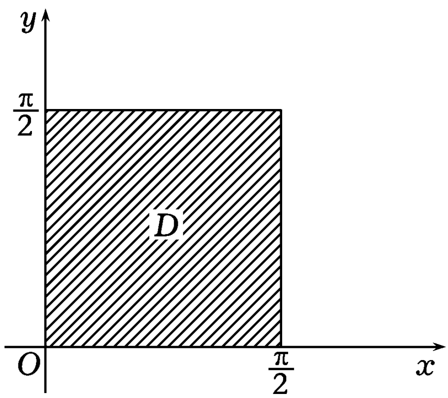


图 3-4

当 $x<0$ 或 $y<0$ 时,$f(x,y)=0$,故

$$F(x, y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\mathrm{d}u\mathrm{d}v=\int_{-\infty}^{x}\int_{-\infty}^{y}0\mathrm{d}u\mathrm{d}v=0;$$

当 $0\leq x<\frac{\pi}{2}, 0\leq y<\frac{\pi}{2}$时,

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\mathrm{d}u\mathrm{d}v=\frac{1}{2}\int_{0}^{x}\int_{0}^{y}\sin(u+v)\mathrm{d}u\mathrm{d}v$$

$$=\frac{1}{2}\int_{0}^{x}\left[-\cos(u+v)\right]_{0}^{y}\mathrm{d}u$$

$$=\frac{1}{2}\int_{0}^{x}[\cos u-\cos(u+y)]\mathrm{d}u$$

$$=\frac{1}{2}[\sin x+\sin y-\sin(x+y)];$$

当 $0\leq x<\frac{\pi}{2}, y\geq\frac{\pi}{2}$时,

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\mathrm{d}u\mathrm{d}v=\frac{1}{2}\int_{0}^{x}\mathrm{d}u\int_{0}^{\frac{\pi}{2}}\sin(u+v)\mathrm{d}v$$

$$=\frac{1}{2}(1+\sin x-\cos x);$$

当 $x \geq \frac{\pi}{2}, 0 \leq y < \frac{\pi}{2}$时,

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\mathrm{d}u\mathrm{d}v = \frac{1}{2}\int_{0}^{\frac{\pi}{2}}\mathrm{d}u\int_{0}^{y}\sin(u + v)\mathrm{d}v$$

$$= \frac{1}{2}(1 + \sin y - \cos y);$$

当 $x \geq \frac{\pi}{2}, y \geq \frac{\pi}{2}$时,

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

$$= \frac{1}{2}\int_{0}^{\frac{\pi}{2}}\mathrm{d}u\int_{0}^{\frac{\pi}{2}}\sin(u + v)\mathrm{d}v = 1.$$

综合上述情形,得

$$F(x, y) = \begin{cases} 0, & x < 0 \text{ 或 } y < 0; \\ \frac{1}{2}[\sin x + \sin y - \sin(x + y)], & 0 \leq x < \frac{\pi}{2}, 0 \leq y < \frac{\pi}{2}; \\ \frac{1}{2}(1 + \sin x - \cos x), & 0 \leq x < \frac{\pi}{2}, y \geq \frac{\pi}{2}; \\ \frac{1}{2}(1 + \sin y - \cos y), & x \geq \frac{\pi}{2}, 0 \leq y < \frac{\pi}{2}; \\ 1, & x \geq \frac{\pi}{2}, y \geq \frac{\pi}{2}. \end{cases}$$

例 3.7 **设**(X, Y)**的分布函数为**

$$F(x, y) = A \quad B + \arctan\frac{x}{2} \quad C + \arctan\frac{y}{3} \quad .$$

求:(1) **系数** A, B, C;

(2) (X, Y)**的联合密度函数**;

(3) X **与** Y **的边缘分布函数和边缘密度函数**;

(4) $P\{-2 < X \leq 2, 0 \leq Y \leq 3\}$**及** $P\{0 \leq X < 2, Y < 3\}$.

解 (1) **由分布函数的性质有**

$$F(+\infty, +\infty) = A \quad B + \frac{\pi}{2} \quad C + \frac{\pi}{2} \quad = 1;$$

$$F(x, +\infty) = A \quad B + \arctan\frac{x}{2} \quad C - \frac{\pi}{2} \quad = 0;$$

$$F(-\infty, y) = A\left(B - \frac{\pi}{2}\right)\left(C + \arctan\frac{y}{3}\right) = 0 .$$

由以上三式解得 $A = \frac{1}{\pi^2}, B = \frac{\pi}{2}, C = \frac{\pi}{2}$,**故**

$$F(x, y) = \frac{1}{\pi^2}\left(\frac{\pi}{2} + \arctan\frac{x}{2}\right)\left(\frac{\pi}{2} + \arctan\frac{y}{3}\right) .$$

(2) (X, Y)**的联合密度函数为**

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{6}{\pi^2 (x^2 + 4)(y^2 + 9)} .$$

(3) X **的边缘分布函数为**

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

$$= \frac{1}{\pi^2}\left(\frac{\pi}{2} + \arctan\frac{x}{2}\right)\left(\frac{\pi}{2} + \frac{\pi}{2}\right)$$

$$= \frac{1}{\pi}\left(\frac{\pi}{2} + \arctan\frac{x}{2}\right)$$

$$= \frac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{2} .$$

边缘密度函数为

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}F_X(x) = \frac{2}{\pi(x^2 + 4)} .$$

类似方法可求得 Y **的边缘分布函数为**

$$F_Y(y) = F(+\infty, y) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{y}{3} .$$

边缘密度函数为

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_Y(y) = \frac{3}{\pi(y^2 + 9)} .$$

(4) **解法一　用联合密度函数** $f(x, y)$**求** .

$$P\{-2 < X \leq 2, 0 \leq Y \leq 3\}$$

$$= \int_{-2}^{2}\int_{0}^{3} f(x, y)\mathrm{d}x\mathrm{d}y = \frac{6}{\pi^2}\int_{-2}^{2}\frac{\mathrm{d}x}{x^2+4}\int_{0}^{3}\frac{\mathrm{d}y}{y^2+9}$$

$$= \frac{1}{8},$$

$$P\{0 \leq X < 2, Y < 3\}$$

$$= \int_{0}^{2}\int_{-\infty}^{3} f(x, y)\mathrm{d}x\mathrm{d}y = \frac{6}{\pi^2}\int_{0}^{2}\frac{\mathrm{d}x}{x^2+4}\int_{-\infty}^{3}\frac{\mathrm{d}y}{y^2+9}$$

$$= \frac{3}{16}.$$

解法二　直接用分布函数 $F(x, y)$**求**.

因　　$P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$

$$= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1).$$

所以　　$P\{-2 < X \leq 2, 0 \leq Y \leq 3\}$

$$= F(2,3) - F(-2,3) - F(2,0) + F(-2,0)$$

$$= \frac{1}{8};$$

$$P\{0 \leq X < 2, Y < 3\}$$

$$= F(2,3) - F(0,3) - F(2, -\infty) + F(0, -\infty)$$

$$= \frac{3}{16}.$$

三、两个随机变量相互独立的判定

例 3.8　**设**(X, Y)**的联合分布列如下**:

$Y \backslash X$	0	1	2	3	$p_{i\cdot}$
0	$\frac{1}{27}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{27}$	$\frac{8}{27}$
1	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	0	$\frac{4}{9}$
2	$\frac{1}{9}$	$\frac{1}{9}$	0	0	$\frac{2}{9}$
3	$\frac{1}{27}$	0	0	0	$\frac{1}{27}$
$p_{\cdot j}$	$\frac{8}{27}$	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{1}{27}$	

判断 X 与 Y 是否相互独立.

解　从(X, Y)的联合分布列可以看出

$$P_{00} = P\{X = 0, Y = 0\} = \frac{1}{27}.$$

从 X 与 Y 的边缘分布列可以看出

$$P_{0\cdot} = P\{X = 0\} = \frac{8}{27}, \quad P_{\cdot 0} = P\{Y = 0\} = \frac{8}{27}.$$

显然

$$P_{00} \neq P_{0\cdot} \cdot P_{\cdot 0},$$

所以 X 与 Y 不相互独立.

例 3.9　设(X, Y)的联合分布列为

(X, Y)	$(1,1)$	$(1,2)$	$(1,3)$	$(2,1)$	$(2,2)$	$(2,3)$
p_{ij}	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{3}$	α	β

问 α, β 取何值时, X, Y 才会相互独立?

解　先由联合分布律求出边缘分布列,写出联合分布列的边缘上(如下表):

$X \backslash Y$	1	2	3	$p_{i\cdot}$
1	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{3}$
2	$\frac{1}{3}$	α	β	$\frac{1}{3}+\alpha+\beta$
$p_{\cdot j}$	$\frac{1}{2}$	$\alpha+\frac{1}{9}$	$\beta+\frac{1}{18}$	1

由于边缘分布必须满足 $\sum_{i=1}^{2} p_{i\cdot} = 1$, $\sum_{j=1}^{3} p_{\cdot j} = 1$, 又 X、Y 相互独立的等价条件为 $p_{ij} = p_{i\cdot}\, p_{\cdot j}(i = 1,2;\ j = 1,2,3)$.因待定参数只有 α, β,故只需列出含 α, β 的两个方程.

$$\begin{cases} \frac{1}{3} + \frac{1}{3} + \alpha + \beta = 1 \\ \frac{1}{9} = \frac{1}{3} \cdot \left(\alpha + \frac{1}{9}\right) \end{cases}, \quad \text{解得} \quad \begin{cases} \alpha = \frac{2}{9} \\ \beta = \frac{1}{9} \end{cases}.$$

经检验,当 $\alpha = \frac{2}{9}$, $\beta = \frac{1}{9}$ 时,对于所有的 $i = 1,2;\ j = 1,2,3$ 均有 $p_{ij} = p_{i\cdot}\, p_{\cdot j}$,所以当 $\alpha = \frac{2}{9}$, $\beta = \frac{1}{9}$ 时, X 与 Y 才会相互独立.

注 对于二维离散型随机向量,若要判定不相互独立,只需找到一个 i, j,使 $p_{ij} \neq p_{\cdot j} \cdot p_{i\cdot}$ 即可;若要判定相互独立,则需要检验所有的 i, j 均满足 $p_{ij} = p_{\cdot j} \cdot p_{i\cdot}$ 才行.如例 3.9 最后的检验是不可缺少的.因为我们为了解方程的简便,只就独立性的部分条件列出了方程,求得的参数值还必须满足独立的全部条件.

例 3.10 设(X, Y)的联合密度函数为

$$f(x, y) = \begin{cases} \frac{1}{2x^2 y}, & 1 \leq x < +\infty, \quad \frac{1}{x} < y < x; \\ 0, & \text{其他}. \end{cases}$$

判断 X 与 Y 是否相互独立.

解 先求 $f_X(x)$ 和 $f_Y(y)$.

当 $x<1$ 时，$f_X(x)=0$;

当 $x\geq 1$ 时，$f_X(x)=\int_{\frac{1}{x}}^{x}\frac{1}{2x^2y}\mathrm{d}y=\frac{1}{2x^2}\ln|y|\Big|_{\frac{1}{x}}^{x}=\frac{1}{x^2}\ln x$;

因此
$$f_X(x)=\begin{cases}\frac{1}{x^2}\ln x & x\geq 1;\\ 0 & x<1.\end{cases}$$

当 $y\leq 0$ 时，$f_Y(y)=0$;

当 $0<y<1$ 时，$f_Y(y)=\int_{\frac{1}{y}}^{+\infty}\frac{1}{2x^2y}\mathrm{d}x=\frac{1}{2}$;

当 $1\leq y<+\infty$ 时，$f_Y(y)=\int_{y}^{+\infty}\frac{1}{2x^2y}\mathrm{d}x=\frac{1}{2y^2}$.

因此
$$f_Y(y)=\begin{cases}\frac{1}{2}, & 0<y<1;\\ \frac{1}{2y^2}, & 1\leq y<+\infty;\\ 0, & \text{其他}.\end{cases}$$

由于
$$f(x,y)\neq f_X(x)\cdot f_Y(y),$$

故 X 与 Y 不相互独立.

例 3.11 一电子仪器由两部件构成，以 X 和 Y 分别表示两部件的寿命(单位：千小时)，已知 X 和 Y 的联合分布函数为
$$F(x,y)=\begin{cases}1-\mathrm{e}^{-x}-\mathrm{e}^{-y}+\mathrm{e}^{-(x+y)}, & x\geq 0, y\geq 0;\\ 0, & \text{其他}.\end{cases}$$

问 X 和 Y 是否相互独立?

解法一 因为 $F_X(x)=F(x,+\infty)=\begin{cases}1-\mathrm{e}^{-x}, & x\geq 0;\\ 0, & x<0.\end{cases}$

$$F_Y(y)=F(+\infty,y)=\begin{cases}1-\mathrm{e}^{-y}, & y\geq 0;\\ 0, & y<0.\end{cases}$$

容易验算当 $x\geq 0, y\geq 0$ 时，有 $F_X(x)\cdot F_Y(y)=F(x,y)$；当 x 和 y 取其他值时，显然有 $F_X(x)\cdot F_Y(y)=F(x,y)=0$，于是对 " $(x,y)\in R^2$，都有 $F(x,y)=F_X(x)\cdot F_Y(y)$，故 X 和 Y 相互独

立.

解法二　因为

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}=\begin{cases}e^{-(x+y)}, & x\geq 0,\ y\geq 0;\\ 0, & \text{其他}.\end{cases}$$

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\begin{cases}\int_0^{+\infty} e^{-(x+y)}\mathrm{d}y, & x\geq 0;\\ 0, & \text{其他}.\end{cases}$$

$$=\begin{cases}e^{-x}, & x\geq 0;\\ 0, & \text{其他}.\end{cases}$$

同理可求得

$$f_Y(y)=\begin{cases}e^{-y}, & y\geq 0;\\ 0, & \text{其他}.\end{cases}$$

所以对" $(x,y)\in R^2$, $f(x,y)=f_X(x)f_Y(y)$成立,于是 X 和 Y 相互独立.

解法三　因为

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}=\begin{cases}e^{-x}\cdot e^{-y}, & x\geq 0,\ y\geq 0;\\ 0, & \text{其他}.\end{cases}$$

即 $f(x,y)=h(x)\cdot g(y)$,且在 x 轴上 $h(x)\geq 0$,在 y 轴上 $g(y)\geq 0$,所以 X 与 Y 相互独立.

四、两个随机变量函数的分布

例 3.12　已知离散型随机向量(X,Y)的分布列为

$X \backslash Y$	0	1	2
0	0.10	0.25	0.15
1	0.15	0.20	0.15

求:(1) $Z=X+Y$;　(2) $Z=XY$;　(3) $Z=\sin\frac{\pi(X+Y)}{2}$;

(4) $Z=\max(X,Y)$的分布列.

解　由(X, Y)的分布列可列表如下

概　　率	0.1	0.25	0.15	0.15	0.20	0.15
(x, y)	$(0,0)$	$(0,1)$	$(0,2)$	$(1,0)$	$(1,1)$	$(1,2)$
$X+Y$	0	1	2	1	2	3
XY	0	0	0	0	1	2
$\sin\frac{\pi}{2}(X+Y)$	0	1	0	1	0	-1
$\max(X, Y)$	0	1	2	1	1	2

在上表中将 Z 取相同值所对应的概率合并,分别得到

(1) $X+Y$ 的分布列为

$X+Y$	0	1	2	3
p_k	0.1	0.25 + 0.15	0.15 + 0.20	0.15

(2) XY 的分布列为

XY	0	1	2
p_k	0.1 + 0.25 + 0.15 + 0.15	0.20	0.15

(3) $\sin\frac{\pi}{2}(X+Y)$的分布列为

$\sin\frac{\pi}{2}(X+Y)$	-1	0	1
p_k	0.15	0.1 + 0.15 + 0.20	0.25 + 0.15

(4) $\max(X, Y)$的分布列为

$\max(X, Y)$	0	1	2
p_k	0.1	0.25 + 0.15 + 0.20	0.15 + 0.15

例 3.13　已知随机变量 X 和 Y 相互独立,且

$$f_X(x)=\begin{cases}1, & 0\leq x\leq 1;\\ 0, & \text{其他}.\end{cases}$$

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

求 $Z=2X+Y$ 的分布函数 $F_Z(z)$.

解 由独立性可知(X,Y)的联合概率密度为

$$f(x,y)=f_X(x)f_Y(y)=\begin{cases}1\cdot e^{-y}, & 0\leq x\leq 1, y>0;\\ 0, & \text{其他}.\end{cases}$$

$$F_Z(z)=P\{Z\leq z\}=P\{2X+Y\leq z\}$$
$$=\iint_{2x+y\leq z} f(x,y)\mathrm{d}x\mathrm{d}y.$$

如图 3-5 所示:

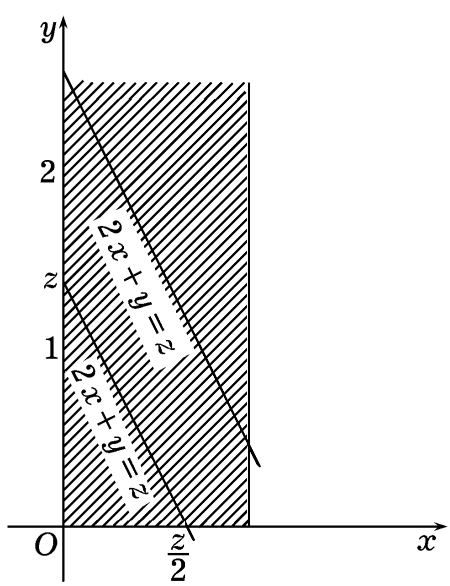


图 3-5

当 $z\leq 0$ 时,

$$F_Z(z)=\iint_{2x+y\leq z} f(x,y)\mathrm{d}x\mathrm{d}y=0;$$

当 $0<z\leq 2$ 时,

$$F_Z(z) = \iint\limits_{2x+y\le z} f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_0^{\frac{z}{2}}\int_0^{z-2x} \mathrm{e}^{-y}\,\mathrm{d}y\mathrm{d}x$$

$$= \int_0^{\frac{z}{2}} (1 - \mathrm{e}^{2x-z})\,\mathrm{d}x = \frac{z}{2} - \frac{1}{2} + \frac{1}{2}\mathrm{e}^{-z};$$

当 $z > 2$ **时**,

$$F_Z(z) = \iint\limits_{2x+y\le z} f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_0^{1}\int_0^{z-2x} \mathrm{e}^{-y}\,\mathrm{d}y\mathrm{d}x.$$

$$= 1 - \frac{1}{2}\mathrm{e}^{2-z} + \frac{1}{2}\mathrm{e}^{-z}.$$

所以

$$F_Z(z) = \begin{cases} 0, & z \le 0; \\ \dfrac{1}{2}(z - 1 + \mathrm{e}^{-z}), & 0 < z \le 2; \\ \dfrac{1}{2}(2 - \mathrm{e}^{2-z} + \mathrm{e}^{-z}), & z > 2. \end{cases}$$

例 3.14　已知随机变量 X, Y **相互独立,均在** $[0,1]$ **上服从均匀分布,求** $Z = X + Y$ **的概率密度函数**.

解　由题设

$$f_X(x) = \begin{cases} 1, & 0 \le x \le 1; \\ 0, & \text{其他}. \end{cases} \qquad f_Y(y) = \begin{cases} 1, & 0 \le y \le 1; \\ 0, & \text{其他}. \end{cases}$$

(X, Y)**的联合概率密度函数为**

$$f(x,y) = f_X(x)f_Y(y) = \begin{cases} 1, & 0 \le x \le 1, 0 \le y \le 1; \\ 0, & \text{其他}. \end{cases}$$

解法一　分布函数法,类似例 3.13,**有**

$$F_Z(z) = P\{Z \le z\} = P\{X + Y \le z\}$$

$$= \iint\limits_{x+y\le z} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

当 $z \le 0$ **时**,

$$F_Z(z) = \iint\limits_{x+y\le z} f(x,y)\,\mathrm{d}x\mathrm{d}y = 0;$$

当 $0 < z \le 1$ **时**,

$$F_Z(z)=\iint\limits_{x+y\le z} f(x,y)\,dx\,dy=\int_0^z\int_0^{z-x} dy\,dx=\frac{z^2}{2};$$

当 $1<z\le 2$ **时**,

$$\begin{aligned}F_Z(z)&=\iint\limits_{x+y\le z} f(x,y)\,dx\,dy\\&=\int_0^{z-1}\int_0^1 dy\,dx+\int_{z-1}^1\int_0^{z-x} dy\,dx\\&=2z-1-\frac{1}{2}z^2;\end{aligned}$$

当 $z>2$ **时**, $F_Z(z)=\iint\limits_{x+y\le z} f(x,y)\,dx\,dy=1$.

因此
$$F_Z(z)=\begin{cases}0, & z\le 0;\\ \dfrac{z^2}{2}, & 0<z\le 1;\\ 2z-1-\dfrac{1}{2}z^2, & 1<z\le 2;\\ 1, & z>2.\end{cases}$$

从而由公式 $f_Z(z)=\dfrac{d}{dz}F_Z(z)$**得到**

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

解法二　直接用卷积公式

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

当 $0\le x\le 1$, $0<z-x\le 1$ **时**, **上式中的被积函数取非零值**, **如图** 3-6 **所示**.

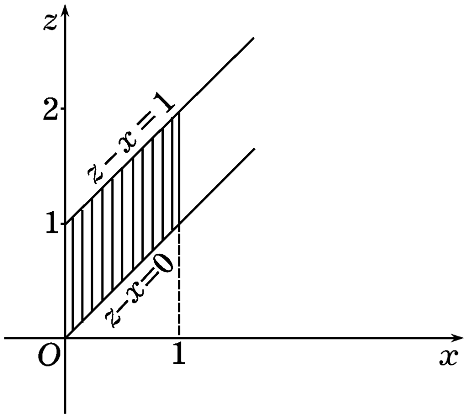


图 3-6

当 $z\le 0$ **时**, $f_Z(z)=0$;

当 $0<z\le 1$ **时**,

$$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x=\int_0^z \mathrm{d}x = z;$$

当 $1<z\leq 2$ 时,

$$f(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x=\int_{z-1}^{1} = 2-z.$$

故得

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

解法三　利用卷积公式再进行变量代换

$$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x=\int_0^1 1\cdot f_Y(z-x)\mathrm{d}x.$$

令 $t=z-x$,则有

$$f_Z(z)=\int_0^1 f_Y(z-x)\mathrm{d}x=\int_z^{z-1} f_Y(t)(-\mathrm{d}t)$$

$$=\int_{z-1}^{z} f_Y(t)\mathrm{d}t.$$

因此,当 $z\leq 0$ 时, $f_Z(z)=\int_{z-1}^{z} f_Y(y)\mathrm{d}y=0;$

当 $0<z\leq 1$ 时,

$$f_Z(z)=\int_{z-1}^{z} f_Y(y)\mathrm{d}y=\int_{z-1}^{0} f_Y(y)\mathrm{d}y+\int_0^z f_Y(y)\mathrm{d}y$$

$$=0+\int_0^z \mathrm{d}y = z;$$

当 $1<z\leq 2$ 时,

$$f_Z(z)=\int_{z-1}^{z} f_Y(y)\mathrm{d}y=\int_{z-1}^{1} f_Y(y)\mathrm{d}y+\int_0^z f_Y(y)\mathrm{d}y$$

$$=\int_{z-1}^{1}\mathrm{d}y+0=2-z;$$

当 $z>2$ 时, $f_Z(z)=\int_{z-1}^{z} f_Y(y)\mathrm{d}y=0.$

故得

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

习 题 选 解

(**3－1**)　**已知二维离散随机变量**(X, Y)**的联合分布列为**

$X \backslash Y$	0	1	2
0	$\frac{1}{4}$	$\frac{1}{6}$	C
1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$

(1) **确定常数** C;

(2) **求**(X, Y)**关于** X, Y **的边缘分布列**.

解　(1) **因为** $\sum_i \sum_j p_{ij} = 1$,**所以有**$\frac{1}{4} + \frac{1}{6} + C + \frac{1}{4} + \frac{1}{8} + \frac{1}{12} = 1$,**求得** $C = \frac{1}{8}$.

(2) $p_{\cdot j} = \sum_i p_{ij}$, $p_{i\cdot} = \sum_j p_{ij}$.

j	0	1	2
$p_{\cdot j}$	$\frac{1}{2}$	$\frac{7}{24}$	$\frac{5}{24}$

i	0	1
$p_{i\cdot}$	$\frac{13}{24}$	$\frac{11}{24}$

(**3－2**)　**箱子里装有** 12 **件产品**,**其中** 2 **件是次品**.**每次从箱子里任取** 1 **件产品**,**共取** 2 **次**.**定义随机变量** X, Y **如下**:

$$X = \begin{cases} 0, & \text{第一次取出正品;} \\ 1, & \text{第一次取出次品.} \end{cases} \qquad Y = \begin{cases} 0, & \text{第二次取出正品;} \\ 1, & \text{第二次取出次品.} \end{cases}$$

分别就下面两种情况求出二维随机向量(X, Y)**的联合分布列和关于** X, Y **的边缘分布列**:

(1) **放回抽样**;

(2) **不放回抽样**.

解 (1) 因为第一次抽样与第二次抽样是相互独立的,所以有

$$p_{00}=P\{X=0,Y=0\}=P\{X=0\}\cdot P\{Y=0\}$$

$$=\frac{10}{12}\times\frac{10}{12}=\frac{25}{36};$$

$$p_{10}=P\{X=1,Y=0\}=P\{X=1\}\cdot P\{Y=0\}$$

$$=\frac{2}{12}\times\frac{10}{12}=\frac{5}{36};$$

同理可求得 $p_{01}=\dfrac{5}{36}$; $p_{11}=\dfrac{1}{36}$.

故(X,Y)的联合分布列及 X,Y 的边缘分布列为

$X\backslash Y$	0	1	$p_{i\cdot}$
0	$\frac{25}{36}$	$\frac{5}{36}$	$\frac{5}{6}$
1	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$p_{\cdot j}$	$\frac{5}{6}$	$\frac{1}{6}$	

(2) 因为第一次抽样与第二次抽样不是相互独立的,所以有

$$p_{00}=P\{X=0,Y=0\}=P\{X=0\}P\{Y=0\mid X=0\}$$

$$=\frac{10}{12}\times\frac{9}{11}=\frac{15}{22};$$

$$p_{10}=P\{X=1,Y=0\}=P\{X=1\}\cdot P\{Y=0\mid X=1\}$$

$$=\frac{2}{12}\times\frac{10}{11}=\frac{15}{33};$$

同理可求得 $p_{01}=\dfrac{10}{12}\times\dfrac{2}{11}=\dfrac{5}{33}$, $p_{11}=\dfrac{2}{12}\times\dfrac{1}{11}=\dfrac{1}{66}$.

故联合分布列及边缘分布列如下

$X\backslash Y$	0	1	$p_{i\cdot}$

0	$\frac{15}{22}$	$\frac{5}{33}$	$\frac{5}{6}$
1	$\frac{5}{33}$	$\frac{1}{66}$	$\frac{1}{6}$
$p_{\cdot j}$	$\frac{5}{6}$	$\frac{1}{6}$	

(3－3)　设二维随机向量(X, Y)的概率密度为

$$f(x, y) = \begin{cases} k\mathrm{e}^{-(3x+4y)}, & x>0, y>0; \\ 0, & \text{其他}. \end{cases}$$

求:(1) **常数** K;

(2) $P\{0\leq X\leq 1, 0\leq Y\leq 2\}$.

解　(1) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\mathrm{d}x\mathrm{d}y$

$$= \int_0^{+\infty}\int_0^{+\infty} k\mathrm{e}^{-3x}\mathrm{e}^{-4y}\mathrm{d}x\mathrm{d}y$$

$$= k\int_0^{+\infty}\mathrm{e}^{-3x}\mathrm{d}x \int_0^{+\infty}\mathrm{e}^{-4y}\mathrm{d}y$$

$$= k\times\frac{1}{3}\times\frac{1}{4} = 1,$$

故 $k = 12$.

(2) $P\{0\leq X\leq 1, 0\leq Y\leq 2\}$

$$= 12\int_0^1\int_0^2 \mathrm{e}^{-(3x+4y)}\mathrm{d}x\mathrm{d}y = 12\int_0^1\mathrm{e}^{-3x}\mathrm{d}x \int_0^2\mathrm{e}^{-4y}\mathrm{d}y$$

$$= (1-\mathrm{e}^{-3})(1-\mathrm{e}^{-8}) \approx 0.949\,9.$$

(3－4)　设随机向量(X, Y)服从二维正态分布 $N(0, 0, 10^2, 10^2, 0)$,**其概率密度为**

$$f(x, y) = \frac{1}{200\pi}\mathrm{e}^{-\frac{x^2+y^2}{200}},$$

求 $P\{x\leq Y\}$.

解　本题若采用直接计算显然比较繁.由图形的对称性可知 $P\{X\leq Y\} = P\{X > Y\}$,**而** $P\{X\leq Y\} + P\{X > Y\} = 1$,**故** $P\{X$

$\leq Y\} = \dfrac{1}{2}$.

(3－5)　设随机向量(X, Y)的概率密度为

$$f(x, y) = \begin{cases} A(R - \sqrt{x^2 + y^2}), & x^2 + y^2 \leq R^2; \\ 0, & x^2 + y^2 > R^2 . \end{cases}$$

求:(1) **常数** A;

(2) (X, Y)**落在圆域** $G: x^2 + y^2 \leq r^2 (r < R)$**中的概率** .

解　(1) **因为**$\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\mathrm{d}x\mathrm{d}y = 1$,

所以
$$\iint\limits_{x^2+y^2<R^2} A(R - \sqrt{x^2 + y^2})\mathrm{d}x\mathrm{d}y$$
$$= \int_0^{2\pi}\mathrm{d}\theta\int_0^R A(R - \rho)\rho\mathrm{d}\rho = 2\pi\int_0^R A(R\rho - \rho^2)\mathrm{d}\rho$$
$$= \frac{\pi AR^3}{3} = 1 .$$

故
$$A = \frac{3}{\pi R^3} .$$

(2)
$$P\{(X, Y) \in G\} = \iint\limits_{x^2+y^2\leq r^2} \frac{3}{\pi R^3}(R - \sqrt{x^2 + y^2})\mathrm{d}x\mathrm{d}y$$
$$= \int_0^{2\pi}\mathrm{d}\theta\int_0^r \frac{3}{\pi R^3}(R - \rho)\rho\mathrm{d}\rho$$
$$= \frac{r^2(3R - 2r)}{R^3} .$$

(3－6)　设二维连续随机向量(X, Y)的概率密度为

$$f(x, y) = \begin{cases} 4xy, & 0\leq x\leq 1, 0\leq y\leq 1; \\ 0, & \text{其他} . \end{cases}$$

求(X, Y)**的分布函数** $F(x, y)$.

解
$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\mathrm{d}u\mathrm{d}v .$$

由于 $f(x, y)$**是分段表达的函数,计算** $F(x, y)$**时要按**(x, y)**的不同位置分区域计算** .

$x<0$ 或 $y<0$ 时，

$$f(x,y)=0,\quad F(x,y)=0;$$

$0\leq x\leq 1, 0\leq y\leq 1$ 时，

$$F(x,y)=\int_0^x\int_0^y 4uv\,\mathrm{d}u\,\mathrm{d}v=x^2y^2;$$

$0\leq x\leq 1, y>1$ 时，

$$F(x,y)=\int_0^x\int_0^1 4uv\,\mathrm{d}u\,\mathrm{d}v=x^2;$$

$x>1, 0\leq y\leq 1$ 时，

$$F(x,y)=\int_0^1\int_0^y 4uv\,\mathrm{d}u\,\mathrm{d}v=y^2;$$

$x>1, y>1$ 时，

$$F(x,y)=\int_0^1\int_0^1 4uv\,\mathrm{d}u\,\mathrm{d}v=1.$$

故

$$F(x,y)=\begin{cases}0, & x<0 \text{ 或 } y<0;\\ x^2y^2, & 0\leq x\leq 1, 0\leq y\leq 1;\\ x^2, & 0\leq x\leq 1, y>1;\\ y^2, & x>1, 0\leq y\leq 1;\\ 1, & x>1, y>1.\end{cases}$$

(3－7)　设二维连续随机向量(X,Y)服从圆域 $G: x^2+y^2\leq R^2$ 上的均匀分布，求关于 X 及关于 Y 的边缘概率密度.

解　(X,Y)的概率密度为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi R^2}, & x^2+y^2\leq R^2;\\ 0, & \text{其他}.\end{cases}$$

当 $-R\leq x\leq R$ 时，

$$f_X(x)=\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}\frac{1}{\pi R^2}\mathrm{d}y=\frac{2\sqrt{R^2-x^2}}{\pi R^2};$$

当$|x|>R$ 时，因 $f(x,y)=0$，故 $f_X(x)=0$，所以有

$$f_X(x)=\begin{cases}\dfrac{2}{\pi R^2}\quad R^2-x^2, & -R\leq x\leq R;\\ 0, & \text{其他}.\end{cases}$$

同理可求得

$$f_Y(y)=\begin{cases}\dfrac{2}{\pi R^2}\quad R^2-y^2, & -R\leq y\leq R;\\ 0, & \text{其他}.\end{cases}$$

(3－8) 设二维连续随机向量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}x^2+\dfrac{xy}{3}, & 0\leq x\leq 1, 0\leq y\leq 2;\\ 0, & \text{其他}.\end{cases}$$

求关于 X 及关于 Y 的边缘概率密度.

解 $$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y.$$

当 $0\leq x\leq 1$ 时,

$$f_X(x)=\int_0^2 x^2+\frac{xy}{3}\ \mathrm{d}y=2x^3+\frac{2}{3}x;$$

当 $x<0$ 或 $x>1$ 时,

因为 $f(x,y)=0$,故 $f_X(x)=0$.

故 $$f_X(x)=\begin{cases}2x^3+\dfrac{2}{3}x, & 0\leq x\leq 1;\\ 0, & \text{其他}.\end{cases}$$

同理 $$f_Y(y)=\begin{cases}\displaystyle\int_0^1 x^2+\frac{xy}{3}\ \mathrm{d}x, & 0\leq y\leq 2;\\ 0, & \text{其他}.\end{cases}$$

$$=\begin{cases}\dfrac{1}{3}+\dfrac{1}{6}y, & 0\leq y\leq 2;\\ 0, & \text{其他}.\end{cases}$$

(3－9) 设二维连续随机向量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}\mathrm{e}^{-y}, & 0<x<y;\\ 0, & \text{其他}.\end{cases}$$

求关于 X 及关于 Y 的边缘概率密度.

解 $$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y.$$

当 $x>0$ 时,

$$f_X(x)=\int_x^{+\infty} \mathrm{e}^{-y}\mathrm{d}y=\mathrm{e}^{-x};$$

当 $x\leq 0$ 时, $f(x,y)=0$, 故 $f_X(x)=0$, 因此

$$f_X(x)=\begin{cases}\mathrm{e}^{-x}, & x>0;\\ 0, & \text{其他}.\end{cases}$$

同理 $$f_Y(y)=\begin{cases}\int_0^y \mathrm{e}^{-y}\mathrm{d}x, & y>0;\\ 0, & \text{其他}.\end{cases}$$

$$=\begin{cases}y\mathrm{e}^{-y}, & y>0;\\ 0, & \text{其他}.\end{cases}$$

(3-10) 设随机向量 (X,Y) 服从二维正态分布, 其概率密度为

$$f(x,y)=\frac{1}{200\pi}\mathrm{e}^{-\frac{x^2+y^2}{200}},$$

证明 X 与 Y 相互独立.

证 $$f_X(x)=\int_{-\infty}^{+\infty}\frac{1}{200\pi}\mathrm{e}^{-\frac{x^2+y^2}{200}}\mathrm{d}y=\frac{1}{200\pi}\mathrm{e}^{-\frac{x^2}{200}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{y^2}{200}}\mathrm{d}y$$

$$=\frac{1}{10\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{200}}\int_{-\infty}^{+\infty}\frac{1}{10\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2\times 10^2}}\mathrm{d}y$$

$$=\frac{1}{10\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{200}}.$$

因为 $$\int_{-\infty}^{+\infty}\frac{1}{10\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2\times 10^2}}\mathrm{d}y=1.$$

同理 $$f_Y(y)=\frac{1}{10\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{200}}.$$

因为 $f(x,y)=f_X(x)f_Y(y)$, 所以 X 与 Y 相互独立.

(3-11) 设随机向量 (X,Y) 服从圆域 $G: x^2+y^2\leq R^2$ 上的

均匀分布,证明 X 与 Y 不独立.

证　因为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi R^2}, & x^2+y^2\leq R^2;\\ 0, & \text{其他}.\end{cases}$$

由(3－7)题可知

$$f_X(x)=\begin{cases}\dfrac{2}{\pi R^2}\quad R^2-x^2, & -R\leq x\leq R;\\ 0, & \text{其他}.\end{cases}$$

$$f_Y(y)=\begin{cases}\dfrac{2}{\pi R^2}\quad R^2-y^2, & -R\leq y\leq R;\\ 0, & \text{其他}.\end{cases}$$

因为 $f(x,y)\neq f_X(x)f_Y(y)$,故 X 与 Y 不相互独立.

(**3－12**)　设 X 与 Y 是相互独立的随机变量,X 服从均匀分布 $u\ \ 0,\frac{1}{5}$,Y 的联合概率密度为

$$f_Y(y)=\begin{cases}5\mathrm{e}^{-5y}, & y>0;\\ 0, & y\leq 0.\end{cases}$$

求随机向量(X,Y)的概率密度.

解　$f_X(x)=\begin{cases}5, & 0\leq x\leq\frac{1}{5};\\ 0, & \text{其他}.\end{cases}$ 由于 x 与 y 相互独立,所以有

$$f(x,y)=f_X(x)f_Y(y)=\begin{cases}25\mathrm{e}^{-5y}, & 0\leq x\leq\frac{1}{5},\ y>0;\\ 0, & \text{其他}.\end{cases}$$

(**3－13**)　某种商品一周需要量是一个随机变量,其概率密度为

$$f(t)=\begin{cases}t\mathrm{e}^{-t}, & t>0;\\ 0, & t\leq 0.\end{cases}$$

且该商品各周的需求量是相互独立的.试求两周、三周需求量的概率密度.

解　设 Z_i 表示第 i 周的需求量$(i=1,2,3)$,它们是独立同分布的随机变量,记两周的需求量为:$Z=Z_1+Z_2$,Z 的概率密度由

$f_Z(t) = \int_{-\infty}^{+\infty} f(u) f(t-u)\mathrm{d}u$ 计算.

当 $u>0$ 且 $t-u>0$ 时,即 $0<u<t$ 时,

$f(u)f(t-u) = u\mathrm{e}^{-u}(t-u)\mathrm{e}^{-(t-u)} = u(t-u)\mathrm{e}^{-t}$;

在其他处,$f(u)f(t-u)=0$.

故
$$f_Z(t) = \begin{cases} \int_0^t u(t-u)\mathrm{e}^{-t}\mathrm{d}u, & t>0; \\ 0, & \text{其他}. \end{cases}$$
$$= \begin{cases} \dfrac{t^3}{6}\mathrm{e}^{-t}, & t>0; \\ 0, & \text{其他}. \end{cases}$$

记三周的需求量为 $W = Z_1 + Z_2 + Z_3 = (Z_1 + Z_2) + Z_3$
$= Z + Z_3$.

$$f_W(t) = \int_{-\infty}^{+\infty} f_Z(u) f(t-u)\mathrm{d}u$$
$$= \begin{cases} \int_0^t \dfrac{u^3\mathrm{e}^{-u}}{6}(t-u)\mathrm{e}^{-(t-u)}\mathrm{d}u, & t>0; \\ 0, & \text{其他}. \end{cases}$$
$$= \begin{cases} \dfrac{t^5}{120}\mathrm{e}^{-t}, & t>0; \\ 0, & \text{其他}. \end{cases}$$

(3-14) 已知 $X\sim t(n)$,证明 $X^2\sim F(1,n)$.

证 因为 $X\sim t(n)$,

所以 X 可表示成 $\dfrac{Z}{\sqrt{Y/n}}$,其中 $Z\sim N(0,1)$, $Y\sim \chi^2(n)$.

因此
$$X^2 = \frac{Z^2}{Y/n} = \frac{Z^2/1}{Y/n}.$$

又因为 $Z\sim N(0,1)$, $Z^2\sim\chi^2(1)$,

按照 F 分布的定义知 $X^2\sim F(1,n)$.

试 题 选 解

[3－1] 是非题

1．设 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X 与 Y 独立，则 $\dfrac{X^2}{Y/n} \sim F(1,n)$．

2．设 $X \sim \chi^2(m)$，$Y \sim \chi^2(n)$，则 $X+Y \sim \chi^2(m+n)$．

3．若已知随机变量 X 与 Y 的分布，则可惟一确定二元随机变量 (X,Y) 的分布．

4．设 $X_1, X_2, \cdots, X_n$ 相互独立，且均服从标准正态分布 $N(0,1)$，则

$$\frac{(n-1)X_1^2}{\sum_{i=1}^{n} X_i^2} \sim F(1,n-1).$$

5．设 $X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，则 $\dfrac{X/n_1}{Y/n_2} \sim F(n_1,n_2)$．

解　1．正确．因为 $X^2 \sim \chi^2(1)$，$Y \sim \chi^2(n)$，所以 $\dfrac{X^2}{Y/n} \sim F(1,n)$．

2．错误．因为 X 与 Y 不一定相互独立．

3．错误．同上题，若 X 与 Y 独立，则结论正确．

4．错误．因为 $X_1^2 \sim \chi^2(1)$，$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$，但 X_1^2 与 $\sum_{i=1}^{n} X_i^2$ 不独立．

5．错误．因为 X 与 Y 不一定相互独立．

[3－2] 填空题

1．设 X 与 Y 独立，各自的密度函数分别为 $f_1(x)$，$f_2(y)$．则 $X+Y$ 的密度函数 $f_{X+Y}(z) =$ ________．

2．设(X, Y)在区域 $D: x^2 + y^2 \leq 4$ 上服从均匀分布，则(X, Y)的联合分布密度函数 $f(x, y) =$ ________．

3．设 $X_1, X_2, \cdots, X_6$ 是来自于 $N(0,1)$ 的样本，则 $C =$ ________时，$C[(X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2]$服从 χ^2 分布．

解　1．$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_1(x) f_2(z - x)\mathrm{d}x$．

2．$f(x, y) = \begin{cases} \dfrac{1}{4\pi}, & x^2 + y^2 \leq 4; \\ 0, & \text{其他}. \end{cases}$

3．$C = \dfrac{1}{3}$.因为 $X_1 + X_2 + X_3 \sim N(0, (\ 3)^2)$，$X_4 + X_5 + X_6 \sim N(0, (\ 3)^2)$，$\dfrac{1}{3}(X_1 + X_2 + X_3) \sim N(0,1)$，$\dfrac{1}{3}(X_4 + X_5 + X_6) \sim N(0,1)$，所以

$$\frac{1}{3}[(X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2] \sim \chi^2(2).$$

[**3－3**]　设随机变量 X, Y 独立，都服从几何分布：

$P = \{X = k\} = P\{Y = k\} = pq^k, k = 0,1,2,\cdots, q = 1 - p$．求 $E[\max(X, Y)]$．

解　先求 $\max[X, Y]$ 的分布律：

$$\begin{aligned}
&P\{\max(X, Y) = k\} \\
&= P\{X \leq k - 1\}P\{Y = k\} + P\{X = k\}P\{Y \leq k - 1\} \\
&\quad + P\{X = k\}P\{Y = k\} \\
&= 2\sum_{i=0}^{k-1} pq^i pq^k + p^2 q^{2k} = 2pq^k - p(2 - p)q^{2k}.
\end{aligned}$$

于是

$$\begin{aligned}
E[\max(X, Y)] &= \sum_{k=0}^{\infty} [2kpq^k - kp(2 - p)q^{2k}] \\
&= 2pq \frac{\mathrm{d}}{\mathrm{d}q} \sum_{k=0}^{\infty} q^k - p(2 - p)q^2
\end{aligned}$$

$$\frac{d}{dq^2}\sum_{k=0}^{\infty} q^{2k}$$

$$= \frac{2(1-p)}{p} - p(2-p)(1-p)^2 \frac{1}{(1-q^2)^2}$$

$$= \frac{(1-p)(3-p)}{p(2-p)}.$$

[3-4] **设随机变量** X **与** Y **独立,且** $P\{X=1\}=P\{X=0\}=P\{X=-1\}=\frac{1}{3}$, Y **服从指数分布:**

$$P\{Y<x\}=\begin{cases}1-e^{-x}, & x>0;\\ 0, & x\leq 0.\end{cases}$$

求 $X+Y$ **的分布密度函数**.

解 $F(x)=P\{X+Y<x\}$

$$=\frac{1}{3}[P\{Y<x+1\}+P\{Y<x\}+P\{Y<x-1\}]$$

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

而 $X+Y$ **的密度函数为** $f(x)=\frac{d}{dx}F(x)$ **即**

$$f(x)=\begin{cases}\frac{1}{3}(e^{-(x+1)}+e^{-x}+e^{-(x-1)}), & x>1;\\ \frac{1}{3}e^{-(x+1)}, & -1<x\leq 0;\\ \frac{1}{3}(e^{-x}+e^{-(x+1)}), & 0<x\leq 1;\\ 0, & x<-1.\end{cases}$$

[3-5] **设随机向量** (X, Y) **的密度函数为**

$$f(x,y)=\begin{cases}\dfrac{C}{(1+x+y)^n}, & x>0,y>0;\\ 0, & \text{其他}.\end{cases}\quad (n>2)$$

求常数 C **及边缘密度函数**.

解　**因为**　$1=\int_0^{+\infty}\int_0^{+\infty}\dfrac{C}{(1+x+y)^n}\mathrm{d}x\mathrm{d}y$

$$=\frac{C}{(n-1)(n-2)}$$

所以

$$C=(n-1)(n-2).$$

当 $x\leq 0$ **时**,　$f_X(x)=0$;

当 $x>0$ **时**,

$$f_X(x)=C\int_0^{+\infty}\frac{1}{(1+x+y)^n}\mathrm{d}y$$

$$=\frac{C}{1-n}\frac{1}{(1+x+y)^{n-1}}\bigg|_0^{+\infty}=\frac{n-2}{(1+x)^{n-1}}$$

$$f_X(x)=\begin{cases}\dfrac{n-2}{(1+x)^{n-1}}, & x>0;\\ 0, & x\leq 0.\end{cases}$$

同理得

$$f_Y(y)=\begin{cases}\dfrac{n-2}{(1+y)^{n-1}}, & y>0;\\ 0, & y\leq 0.\end{cases}$$

[**3－6**]　**设随机变量** X,Y **独立,且** $P\{X=0\}=P\{X=1\}=\frac{1}{2}$, $P\{Y<x\}=x, 0\leq x\leq 1$, **求** $X+Y$ **的密度函数**.

解　**由于**

$$P\{X+Y\leq x\}=P\{X+Y\leq x\mid X=0\}P\{X=0\}+P\{X+Y\leq x\mid X=1\}P\{X=1\}$$

$$=\frac{1}{2}P\{Y\leq x\}+\frac{1}{2}P\{Y\leq x-1\}$$

$$= \begin{cases} \frac{1}{2}x, & 0 \le x \le 1; \\ \frac{1}{2} + \frac{x-1}{2}, & 1 < x \le 2; \\ 1, & x > 2. \end{cases}$$

从而密度函数为

$$f_{X+Y}(x) = \begin{cases} \frac{1}{2}, & 0 \le x \le 2; \\ 0, & \text{其他}. \end{cases}$$

[**3 - 7**] **设** (X, Y) **的联合密度函数为**

$$f(x, y) = \begin{cases} Ce^{-(x+y)}, & x \ge 0, y \ge 0; \\ 0, & \text{其他}. \end{cases}$$

(1) **求常数** C;

(2) **求** $P\{0 \le X \le 1, 0 \le Y \le 1\}$;

(3) X **与** Y **是否相互独立?为什么?**

解 (1) **因为** $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\mathrm{d}x\mathrm{d}y = 1$, **所以**

$$1 = \int_0^{+\infty}\int_0^{+\infty} Ce^{-(x+y)}\mathrm{d}x\mathrm{d}y = C\int_0^{+\infty} e^{-x}\mathrm{d}x\int_0^{+\infty} e^{-y}\mathrm{d}y = C,$$

即 $$C = 1.$$

(2) $$P\{0 \le X \le 1, 0 \le Y \le 1\}$$

$$= \int_0^1\int_0^1 e^{-(x+y)}\mathrm{d}x\mathrm{d}y = \left(1 - \frac{1}{e}\right)^2.$$

(3) $$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\mathrm{d}y = \begin{cases} \int_0^{+\infty} e^{-(x+y)}\mathrm{d}y, & x \ge 0; \\ 0, & \text{其他}. \end{cases}$$

即 $$f_X(x) = \begin{cases} e^{-x}, & x \ge 0; \\ 0, & \text{其他}. \end{cases}$$

同理 $$f_Y(y) = \begin{cases} e^{-y}, & y \ge 0; \\ 0, & \text{其他}. \end{cases}$$

显然 $f_X(x) \cdot f_Y(y) = f(x, y)$, **故** X **与** Y **相互独立**.

[3-8] 设(X,Y)服从圆域 $X^2+Y^2\leq 9$ 上的均匀分布.

(1) 求 $P\{X^2+Y^2\leq 4\}$;

(2) X 与 Y 是否相互独立?为什么?

解 由题设(X,Y)的密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{9\pi}, & x^2+y^2\leq 9;\\ 0, & \text{其他}.\end{cases}$$

(1) $$P\{X^2+Y^2\leq 4\}=\iint\limits_{x^2+y^2\leq 4} f(x,y)\,\mathrm{d}x\mathrm{d}y=\frac{1}{9\pi}\iint\limits_{x^2+y^2\leq 4}\mathrm{d}x\mathrm{d}y=\frac{4}{9}.$$

(2) $$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}y$$

$$=\begin{cases}\dfrac{1}{9\pi}\displaystyle\int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}}\mathrm{d}y, & -3\leq x\leq 3;\\ 0, & \text{其他}.\end{cases}$$

即 $$f_X(x)=\begin{cases}\dfrac{2}{9\pi}\sqrt{9-x^2}, & -3\leq x\leq 3;\\ 0, & \text{其他}.\end{cases}$$

同理可得 $$f_Y(y)=\begin{cases}\dfrac{2}{9\pi}\sqrt{9-y^2}, & -3\leq y\leq 3;\\ 0, & \text{其他}.\end{cases}$$

由于 $f_X(x)f_Y(y)\neq f(x,y)$,故 X 与 Y 不独立.

[3-9] 设二维随机向量(X,Y)的联合概率密度为

$$f(x,y)=\begin{cases}\mathrm{e}^{-y}, & 0<x<y;\\ 0, & \text{其他}.\end{cases}$$

(1) 求(X,Y)分别关于 X 和 Y 的边缘概率密度 $f_X(x)$, $f_Y(y)$.

(2) 判断 X 与 Y 是否相互独立?并说明理由.

(3) 计算 $P\{X+Y\leq 1\}$.

解　(1) $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$

$$=\begin{cases}\int_x^{+\infty}\mathrm{e}^{-y}\mathrm{d}y=\mathrm{e}^{x}, & x>0;\\ 0, & \text{其他}.\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x=\begin{cases}\int_0^y\mathrm{e}^{-y}\mathrm{d}x=y\mathrm{e}^{-y}, & y>0;\\ 0, & \text{其他}.\end{cases}$$

(2) 由于 $f(x,y)\neq f_X(x)\cdot f_Y(y)$，故 X 与 Y 不独立.

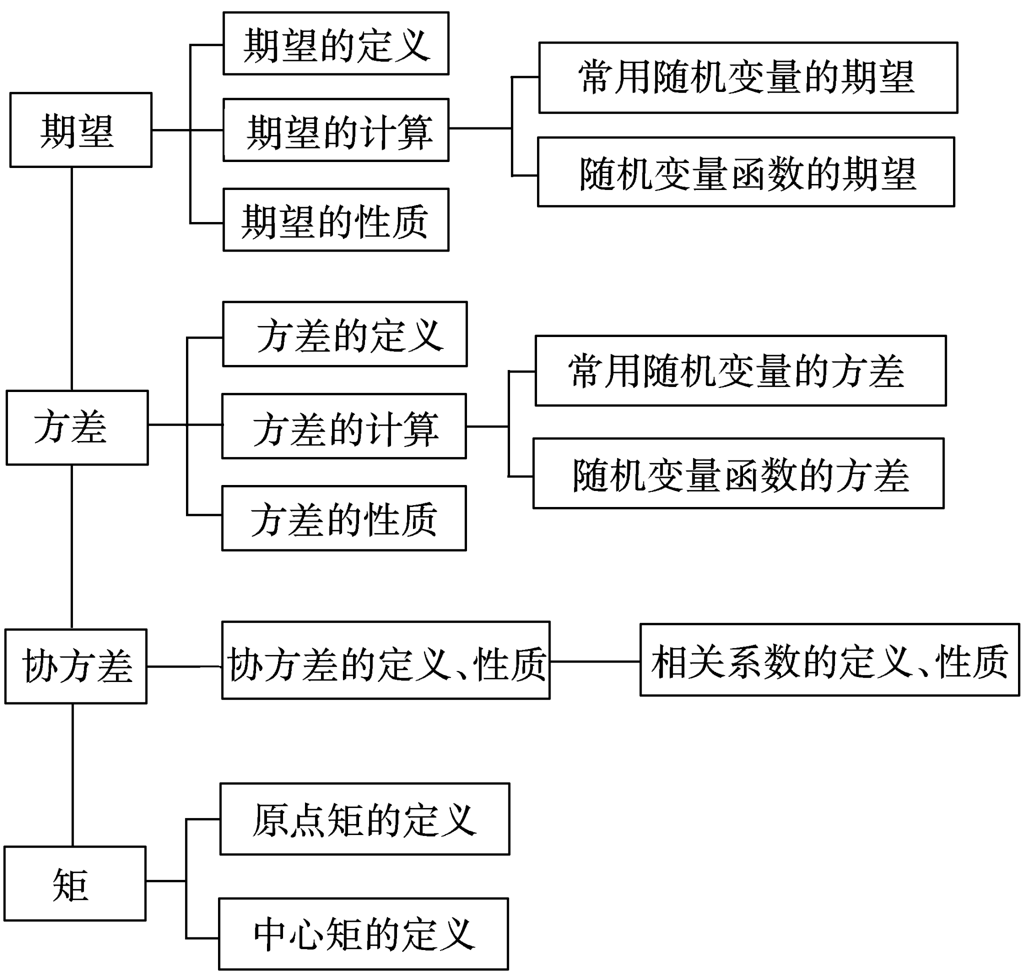
(3) $P\{X+Y\leq 1\}=\iint\limits_{x+y\leq 1}f(x,y)\mathrm{d}x\mathrm{d}y=\int_0^{\frac{1}{2}}\mathrm{d}x\int_x^{1-x}\mathrm{e}^{-y}\mathrm{d}y$

$$=1+\mathrm{e}^{-1}-2\mathrm{e}^{-\frac{1}{2}}.$$

第四章　随机变量的数字特征

内 容 提 要

一、知识脉络



二、知识要点

(一) 随机变量的数学期望

(1) 设离散型随机变量 X 的分布律为 $P\{X=x_k\}=p_k,1,2,\cdots$,若级数 $\sum_k x_k p_k$ 绝对收敛,则称

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

为随机变量 X 的数学期望.

(2) 设连续型随机变量 X 的密度函数为 $f(x)$,若积分 $\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$ 绝对收敛,则称

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

为随机变量 X 的数学期望.

(二) 随机变量函数的数学期望

设 $y=g(x)$为连续函数,$Y=g(X)$为随机变量 X 的函数.

(1) 离散型随机变量 X 的分布列为 $P\{X=x_k\}=p_k,k=1,2,\cdots$,若级数 $\sum_k g(x_k)p_k$ 绝对收敛,则

$$E(Y)=E[g(x)]=\sum_k g(x_k)p_k.$$

(2) 连续型随机变量 X 的密度函数为 $f(x)$,若积分 $\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$ 绝对收敛,则

$$E(Y)=E[g(x)]=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x.$$

设 $z=g(x,y)$是二元连续函数,$Z=g(X,Y)$是二维随机变量(X,Y)的函数.

(3) 若二维离散型随机向量(X,Y)的分布列为

$$P\{X=x_i,Y=y_j\}=p_{ij},\quad i,j=1,2,\cdots,$$

且 $\sum_i \sum_j g(x_i, y_j) p_{ij}$ 绝对收敛,则

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

(4) 若二维连续型随机向量 (X, Y) 的密度函数为 $f(x, y)$,且 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y) f(x, y) \mathrm{d}x \mathrm{d}y$ 绝对收敛,则

$$E(Z) = E[g(X, Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y) f(x, y) \mathrm{d}x \mathrm{d}y .$$

(三) 数学期望的性质

(1) a 为常数, $E(a) = a$.

(2) $E(aX) = aE(X)$, a 为常数 .

(3) $E(aX + b) = aE(X) + b$, a, b 为常数 .

(4) $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$.

(5) 若随机变量 $X_1, X_2, \cdots, X_n$ 相互独立,则

$$E(X_1 X_2 \cdots X_n) = E(X_1) E(X_2) \cdots E(X_n) .$$

(四) 方差

1 . 定义

设 $E(X)$ 为随机变量 X 的数学期望,若 $E[X - E(X)]^2$ 存在,则称 $D(X) = E[X - E(X)]^2$ 为 X 的方差,而 $\sqrt{D(X)}$ 称为 X 的标准差 .

若 X 为离散型,则

$$D(X) = \sum_k (x_k - E(X))^2 p_k ;$$

若 X 为连续型,则

$$D(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) \mathrm{d}x .$$

2 . 方差的另一个计算公式

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

3 . 方差的性质

(1) $D(a) = 0$, a 为常数 .

(2) $D(aX + b) = a^2 D(X)$, $a \neq 0$, b 为常数 .

(3) 若 $X_1, X_2, \cdots, X_n$ 相互独立,则有

$$D(X_1 + X_2 + \cdots + X_n) = D(X_1) + D(X_2) + \cdots + D(X_n).$$

(4) $D(X \pm Y) = D(X) + D(Y) \pm 2\mathrm{cov}(X, Y)$.

(五) 协方差与相关系数

1. 定义

设 X, Y **是两个随机变量**, $D(X)$, $D(Y)$ **存在,则称**
$\mathrm{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$
为 X **与** Y **的协方差,称**

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{D(X) \quad D(Y)} \quad (D(X) > 0, D(Y) > 0)$$

为 X **与** Y **的相关系数或标准协方差**.

2. 协方差性质

(1) $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$; $\mathrm{cov}(X, X) = D(X)$.

(2) $\mathrm{cov}(aX, bY) = ab\mathrm{cov}(X, Y)$, a, b **为常数**.

(3) $\mathrm{cov}(X_1 + X_2, Y) = \mathrm{cov}(X_1, Y) + \mathrm{cov}(X_2, Y)$.

3. 相关系数 ρ_{XY} **性质**

(1) $|\rho_{XY}| \leq 1$.

(2) $|\rho_{XY}| = 1$ **存在** a, b **且** $a \neq 0$ **使** $P\{Y = aX + b\} = 1$.

当 $0 < \rho_{XY} < 1$ **时,称** X **与** Y **正相关**;

当 $-1 < \rho_{XY} < 0$ **时,称** X **与** Y **负相关**;

当 $\rho_{XY} = 0$ **时,称** X **与** Y **不相关**.

(3) **若** X **与** Y **独立,则** X **与** Y **一定不相关,反之不一定真**.
但对二维正态随机变量 (X, Y), X **与** Y **独立** X **与** Y **不相关**.

要 点 范 例

一、求离散型随机变量的期望与方差

例 4.1 **一批零件中有 9 件合格品和 3 件废品,安装机器时**

从这批零件中任取 1 件,如果取出的废品不再放回去,求在取得合格品以前已取出的废品数的数学期望.

解 以 X 表示在取得合格品以前取出的废品个数.X 是一随机变量,它的可能值是 0,1,2,3.取这些值的概率分别为

$$P\{X=0\}=\frac{9}{12}=\frac{3}{4},$$

$$P\{X=1\}=\frac{3}{12}\times\frac{9}{11}=\frac{9}{44},$$

$$P\{X=2\}=\frac{3}{12}\times\frac{2}{11}\times\frac{9}{10}=\frac{9}{220},$$

$$P\{X=3\}=\frac{3}{12}\times\frac{2}{11}\times\frac{1}{10}\times\frac{9}{9}=\frac{1}{220}.$$

所以
$$E(X)=\sum_{k=0}^{3}kp\{X=k\}$$
$$=0\times\frac{3}{4}+1\times\frac{9}{44}+2\times\frac{9}{220}+3\times\frac{1}{220}=\frac{3}{10}.$$

例 4.2 已知(X,Y)的分布列为

X \\ Y	-1	1	2
-1	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{5}{20}$
2	$\frac{2}{20}$	$\frac{6}{20}$	$\frac{3}{20}$

试求:(1) $E(X),D(X);E(Y),D(Y)$;

(2) $E(X-Y),D(X-Y)$;

(3) $E(XY),D(XY)$.

解 先求出关于 X 和关于 Y 的边缘分布列,

X	-1	2
$p_{i\cdot}$	$\frac{9}{20}$	$\frac{11}{20}$

Y	-1	1	2
$p_{\cdot j}$	$\frac{5}{20}$	$\frac{7}{20}$	$\frac{8}{20}$

(1) $E(X)=(-1)\times\frac{9}{20}+2\times\frac{11}{20}=\frac{13}{20}$,

$E(X^2)=(-1)^2\times\frac{9}{20}+2^2\times\frac{11}{20}=\frac{53}{20}$,

$D(X)=E(X^2)-[E(X)]^2=\frac{53}{20}-\left(\frac{13}{20}\right)^2=\frac{891}{400}$.

$E(Y)=(-1)\times\frac{5}{20}+1\times\frac{7}{20}+2\times\frac{8}{20}=\frac{18}{20}$,

$E(Y^2)=(-1)^2\times\frac{5}{20}+1^2\times\frac{7}{20}+2^2\times\frac{8}{20}=\frac{44}{20}$,

$D(Y)=E(Y^2)-[E(Y)]^2=\frac{44}{20}-\left(\frac{18}{20}\right)^2=\frac{556}{400}$.

在解(2)、(3)**之前,求出** $X-Y$, XY **的分布列如下**.

p_k	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{5}{20}$	$\frac{2}{20}$	$\frac{6}{20}$	$\frac{3}{20}$
(X,Y)	$(-1,-1)$	$(-1,1)$	$(-1,2)$	$(2,-1)$	$(2,1)$	$(2,2)$
$X-Y$	0	-2	-3	3	1	0
XY	1	-1	-2	-2	2	4

(2) $X-Y$ **的分布列为**

$X-Y$	-3	-2	0	1	3
p_k	$\frac{5}{20}$	$\frac{1}{20}$	$\frac{6}{20}$	$\frac{6}{20}$	$\frac{2}{20}$

$$\begin{aligned}E(X-Y)&=(-3)\times\frac{5}{20}+(-2)\times\frac{1}{20}+0\times\frac{6}{20}\\&\quad+1\times\frac{6}{20}+3\times\frac{2}{20}\\&=-\frac{5}{20}.\end{aligned}$$

(**此处也可按** $E(X-Y)=E(X)-E(Y)=\frac{13}{20}-\frac{18}{20}=-\frac{5}{20}$**计算**,**而** $D(X-Y)\neq D(X)+(-1)^2D(Y)$,**这是因为** X, Y **不独立**.)

$$E[(X-Y)^2]=(-3)^2\times\frac{5}{20}+(-2)^2\times\frac{1}{20}+0^2\times\frac{6}{20}$$

$$+1^2\times\frac{6}{20}+3^2\times\frac{2}{20}$$

$$=\frac{63}{20},$$

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

$$=\frac{63}{20}-\left(-\frac{5}{20}\right)^2=\frac{1\,235}{400}.$$

(3) XY 的分布列为

XY	-2	-1	1	2	4
p_k	$\frac{7}{20}$	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{6}{20}$	$\frac{3}{20}$

$$E(XY)=(-1)\times\frac{1}{20}+(-2)\times\frac{7}{20}$$

$$+1\times\frac{3}{20}+2\times\frac{6}{20}+4\times\frac{3}{20}$$

$$=\frac{12}{20},$$

$$E[(XY)^2]=(-1)^2\times\frac{1}{20}+(-2)^2\times\frac{7}{20}$$

$$+1^2\times\frac{3}{20}+2^2\times\frac{6}{20}+4^2\times\frac{3}{20}$$

$$=\frac{104}{20},$$

$$D(XY)=E[(XY)^2]-[E(XY)]^2$$

$$=\frac{104}{20}-\left(\frac{12}{20}\right)^2=\frac{1\,936}{400}.$$

例 4.3　已知离散型随机变量 X 的分布函数为

$$F(X)=\begin{cases}0, & x<-1;\\ 0.4, & -1\leq x<0;\\ 0.7, & 0\leq x<1;\\ 1, & x\geq 1.\end{cases}$$

求:(1) $E(X)$;　(2) $D(X)$;　(3) $E(3X^2+5)$.

解　首先由分布函数 $F(x)$ 求出 X 的分布列.X 的可能值为

$F(x)$的跳跃间断点：$-1,0,1$.而 X 取各可能值的概率为 $F(x)$在 X 各可能值处的跳跃值.

$$P\{X=-1\}=F(-1)-F(-1-0)=0.4-0=0.4,$$
$$P\{X=0\}=F(0)-F(0-0)=0.7-0.4=0.3,$$
$$P\{X=1\}=F(1)-F(1-0)=1-0.7=0.3.$$

(此处 $F(-1-0)$表示 $F(x)$在 $x=-1$ 处的左极限,其余同.)即 X 的分布列为

X	-1	0	1
p_k	0.4	0.3	0.3

$$E(X)=(-1)\times 0.4+0\times 0.3+1\times 0.3=-0.1,$$
$$E(X^2)=(-1)^2\times 0.4+0^2\times 0.3+1^2\times 0.3=0.7,$$
$$D(X)=E(X^2)-[E(X)]^2=0.7-(-0.1)^2=0.69,$$
$$E(3X^2+5)=3E(X^2)+5=3\times 0.7+5=7.1.$$

例 4.4 设某年龄段一位健康者(一般体检未发现病症),在 10 年内活着或自杀死亡的概率为 $p(0<p<1$ 已知),在 10 年内非自杀死亡的概率为 $1-p$.保险公司开办 10 年人寿保险,参加者需要交保险费 a 元(已知).若 10 年内非自杀死亡,公司赔偿 b 元 $(b>a)$, b 应如何定才能使公司期望获益;若有 m 人参加保险,公司可望从中收益多少?

解 设 X_k 表示公司从第 k 个参保者身上所得收益,其分布列为

X_k	a	$a-b$
p_k	p	$1-p$

公司期望收益 $E(X_k)>0$,而

$$E(X_k)=ap+(a-b)(1-p)=a-b(1-p),$$

则有 $a>b(1-p)$,结合 $b>a$,即有

$$a<b<\frac{a}{1-p}.$$

若有 m 人参加保险,公司收益为 $X=\sum_{k=1}^{m}X_k$,则

$$E(X)=\sum_{k=1}^{m}E(X_k)=ma-mb(1-p).$$

二、求连续型随机变量的期望与方差

例 4.5 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 $E(X),D(X)$.

解 因为 X 是连续型随机变量,$F(x)$是连续函数,所以

$$0=F(-1-0)=F(-1)=a+b\arcsin(-1)=a-\frac{\pi}{2}b,$$

$$1=F(1-0)=F(1)=a+b\arcsin 1=a+\frac{\pi}{2}b.$$

解得

$$a=\frac{1}{2},\quad b=\frac{1}{\pi}.$$

于是

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

对 $F(x)$求导,得 X 的密度函数

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

所以

$$E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\frac{1}{\pi}\int_{-1}^{1}x\cdot\frac{1}{\sqrt{1-x^2}}\mathrm{d}x=0;$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x=\frac{1}{\pi}\int_{-1}^{1}x^2\cdot\frac{1}{\sqrt{1-x^2}}\mathrm{d}x$$

$$\stackrel{\text{令 } x = \sin t}{=} \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^2 t}{\cos t}\cos t\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}(1-\cos 2t)\,\mathrm{d}t = \frac{1}{2};$$

$$D(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - 0 = \frac{1}{2}.$$

例 4.6　设某种商品每周的需求量 X 是服从区间 $[10, 30]$ 上均匀分布的随机变量, 而经销商店进货量为区间 $[10, 30]$ 中的某一整数, 商店每销售一单位商品可获利 500 元, 若供大于求, 则削价处理, 每处理一单位商品亏损 100 元, 若供不应求, 则可从外部调剂供应, 此时一单位商品仅获利 300 元. 为使商品所获利润期望值不少于 9 280 元, 试确定最少进货量.

解　设进货量为 a, 用 X 表示商品的需求量, 则利润为

$$Y_a = \begin{cases} 500a + (X - a)300, & a < X \leqslant 30; \\ 500X - (a - X)100, & 10 \leqslant X \leqslant a. \end{cases}$$

$$= \begin{cases} 300X + 200a, & a < X \leqslant 30; \\ 600X - 100a, & 10 \leqslant X \leqslant a. \end{cases}$$

期望利润为

$$E(Y_a) = \int_{10}^{30} \frac{1}{20} Y_a \,\mathrm{d}x$$

$$= \frac{1}{20}\int_{10}^{a}(600x - 100a)\,\mathrm{d}x + \int_{a}^{30}(300x + 200a)\,\mathrm{d}x$$

$$= -7.5a^2 + 350a + 5\,250.$$

依题意有

$$-7.5a^2 + 350a + 5\,250 \geqslant 9\,280,$$

即
$$7.5a^2 - 350a + 4\,030 \leqslant 0,$$

解得
$$20\frac{2}{3} \leqslant a \leqslant 26.$$

故利润期望值不少于 9 280 元的最少进货量为 21 单位.

例 4.7　设 (X, Y) 的联合概率密度函数为

$$f(x,y)=\begin{cases}\frac{1}{2}\sin(x+y), & (x,y)\in D;\\ 0, & \text{其他}.\end{cases}$$

其中 D 为矩形域 $0\leq x\leq\frac{\pi}{2}, 0\leq y\leq\frac{\pi}{2}$,试求:

(1) $E(X), E(Y), D(X), D(Y)$; (2) $E(XY)$.

解法一 联合密度法.

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\,\mathrm{d}x\mathrm{d}y=\iint_D x\cdot\frac{1}{2}\sin(x+y)\,\mathrm{d}x\mathrm{d}y$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}\mathrm{d}x\int_0^{\frac{\pi}{2}}x\sin(x+y)\,\mathrm{d}y$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}\left[-x\cos(x+y)\right]_0^{\frac{\pi}{2}}\mathrm{d}x$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}x(\sin x+\cos x)\,\mathrm{d}x=\frac{\pi}{4}.$$

$$E(X^2)=\frac{1}{2}\iint_D x^2 f(x,y)\,\mathrm{d}x\mathrm{d}y=\frac{1}{2}\int_0^{\frac{\pi}{2}}\mathrm{d}x\int_0^{\frac{\pi}{2}}x^2\sin(x+y)\,\mathrm{d}y$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}\left[-x^2\cos(x+y)\right]_0^{\frac{\pi}{2}}\mathrm{d}x$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}x^2(\sin x+\cos x)\,\mathrm{d}x=\frac{\pi^2}{8}+\frac{\pi}{2}-2.$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{\pi^2}{16}+\frac{\pi}{2}-2.$$

由 $f(x,y)$的表达式知, X 与 Y 有对称关系,故

$$E(Y)=\frac{\pi}{4},\quad D(Y)=\frac{\pi^2}{16}+\frac{\pi}{2}-2.$$

解法二 边缘密度法.

先求出 $f_X(x)$.

当 $x<0$ 或 $x>\frac{\pi}{2}$时, $f_X(x)=0$;

当 $0\leq x\leq\frac{\pi}{2}$时,

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\frac{1}{2}\int_0^{\frac{\pi}{2}}\sin(x+y)\mathrm{d}y$$

$$=\frac{1}{2}(\sin x+\cos x).$$

故

$$f_X(x)=\begin{cases}\frac{1}{2}(\sin x+\cos x), & 0\leq x\leq\frac{\pi}{2};\\ 0, & \text{其他}.\end{cases}$$

于是

$$E(X)=\int_{-\infty}^{+\infty}f_X(x)\mathrm{d}x$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}x(\sin x+\cos x)\mathrm{d}x=\frac{\pi}{4};$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2f_X(x)\mathrm{d}x=\frac{1}{2}\int_0^{\frac{\pi}{2}}x^2(\sin x+\cos x)\mathrm{d}x$$

$$=\frac{\pi^2}{8}+\frac{\pi}{2}-2.$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{\pi^2}{16}+\frac{\pi}{2}-2.$$

由 X 与 Y 的对称关系,知

$$E(Y)=\frac{\pi}{4},\quad D(Y)=\frac{\pi^2}{16}+\frac{\pi}{2}-2.$$

(2)

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)\mathrm{d}x\mathrm{d}y$$

$$=\frac{1}{2}\iint_D xy\sin(x+y)\mathrm{d}x\mathrm{d}y$$

$$=\frac{1}{2}\int_0^{\frac{\pi}{2}}\mathrm{d}x\int_0^{\frac{\pi}{2}}xy\sin(x+y)\mathrm{d}y=\frac{\pi}{2}-1.$$

例 4.8 在长为 a 的线段上任取两点,求两点间距离的数学期望和方差.

解 把线段置于数轴上,使它与区间$[0,a]$重合.设 X, Y 分

别表示任取两点的坐标,则 X 与 Y 相互独立,且都在 $[0, a]$ 上服从均匀分布,故 (X, Y) 的联合密度为

$$f(x, y) = f_X(x) f_Y(y) = \begin{cases} \dfrac{1}{a^2}, & 0 \le x \le a, 0 \le y \le a; \\ 0, & \text{其他}. \end{cases}$$

于是两点间的距离 $|X - Y|$ 的期望为

$$\begin{aligned} E(|X - Y|) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |x - y| f(x, y) \mathrm{d}x\mathrm{d}y \\ &= \int_0^a\int_0^a |x - y| \cdot \frac{1}{a^2} \mathrm{d}x\mathrm{d}y \\ &= \frac{1}{a^2}\left[\int_0^a \mathrm{d}x \int_0^x (x - y)\mathrm{d}y + \int_0^a \mathrm{d}x \int_x^a (y - x)\mathrm{d}y\right] \\ &= \frac{a}{3}; \end{aligned}$$

而
$$\begin{aligned} E(|X - Y|^2) &= \int_0^a\int_0^a |x - y|^2 \cdot \frac{1}{a^2} \mathrm{d}x\mathrm{d}y \\ &= \frac{1}{a^2}\int_0^a \mathrm{d}x \int_0^a (x - y)^2 \mathrm{d}y = \frac{a^2}{6}; \end{aligned}$$

故两点间的距离 $|X - Y|$ 的方差为

$$\begin{aligned} D(|X - Y|) &= E(|X - Y|^2) - [E(|X - Y|)]^2 \\ &= \frac{a^2}{6} - \left(\frac{a}{3}\right)^2 = \frac{a^2}{18}. \end{aligned}$$

三、协方差与相关系数的计算

例 4.9 设离散型随机变量 (X, Y) 的联合分布列为

X \ Y	-1	0	1
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

0	$\frac{1}{8}$	0	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

(1) 求 $\operatorname{cov}(X, Y)$, ρ_{XY} .

(2) X 和 Y 是否相关 ? 是否独立 ? 为什么 ?

解 (1) 先由联合分布列分别求得关于 X, Y 的边缘分布律和函数 XY 的分布律, 它们分别为

X	-1	0	1
$p_{i\cdot}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{3}{8}$

Y	-1	0	1
$p_{\cdot j}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{3}{8}$

XY	-1	0	1
p_k	$\frac{2}{8}$	$\frac{4}{8}$	$\frac{2}{8}$

故
$$(X) = E(Y) = (-1)\times\frac{3}{8} + 0\times\frac{2}{8} + 1\times\frac{3}{8} = 0;$$
$$E(XY) = (-1)\times\frac{2}{8} + 0\times\frac{4}{8} + 1\times\frac{2}{8} = 0;$$

从而
$$\operatorname{cov}(X, Y) = E(XY) - E(X)E(Y) = 0;$$
$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{D(X)\quad D(Y)} = 0 .$$

(2) 因 $\rho_{XY} = 0$, 所以 X 和 Y 不相关 . 而联合分布律中 $p_{22} = 0$, $p_{2\cdot}\cdot p_{\cdot 2} = \frac{2}{8}\cdot\frac{2}{8} = \frac{1}{16}$, $p_{22} \neq p_{\cdot 2}\cdot p_{2}$.

故 X 与 Y 不相互独立 .

例 4 .10 设随机变量 X 服从拉普拉斯分布, 其概率密度为

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

(1) 求 X 与$|X|$的协方差和相关系数 .

(2) 问 X 和$|X|$是否相关 ?是否独立 ?为什么 ?

解 (1) $E(X)=\int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x=\int_{-\infty}^{+\infty} x\cdot\frac{1}{2}\mathrm{e}^{-|x|}\,\mathrm{d}x=0$,

$$E(|X|)=\frac{1}{2}\int_{-\infty}^{+\infty}|x|\mathrm{e}^{-|x|}\,\mathrm{d}x=\int_{0}^{+\infty} x\mathrm{e}^{-x}\,\mathrm{d}x=1 .$$

$$\begin{aligned}E(X|X|)&=\int_{-\infty}^{+\infty} x|x|\cdot\frac{1}{2}\mathrm{e}^{-|x|}\,\mathrm{d}x\\&=\frac{1}{2}\int_{-\infty}^{0} x^2\mathrm{e}^{x}\,\mathrm{d}x+\frac{1}{2}\int_{0}^{+\infty} x^2\mathrm{e}^{-x}\,\mathrm{d}x\\&=-\frac{1}{2}+\frac{1}{2}=0 .\end{aligned}$$

于是 $\mathrm{cov}(X,Y)=E(X|X|)-E(X)E(|X|)=0$,从而 $\rho_{X|X|}=0$.

(2) 因 $\rho_{X|X|}=0$,所以 X 和$|X|$不相关 .但 X 与$|X|$不相互独立 .事实上,$P\{X\leq -3,|X|\leq 1\}=P\{$ě$\}=0$,而

$$P\{X\leq -3\}\cdot P\{|X|\leq 1\}=\int_{-\infty}^{-3}\frac{1}{2}\mathrm{e}^{-|x|}\,\mathrm{d}x\cdot\int_{-1}^{1}\frac{1}{2}\mathrm{e}^{-|x|}\,\mathrm{d}x\neq 0,$$

所以

$$P\{X\leq -3\}\cdot P\{|X|\leq 1\}\neq P\{X\leq -3,|X|\leq 1\},$$

于是 X 与$|X|$不相互独立 .

注 以上两例告诉我们,随机变量 X 和 Y 不相关(即 $\rho(X,Y)=0$),并不保证 X,Y 相互独立 .一般而言,只有当(X,Y)是联合正态分布时,才有 X,Y 不相关等价 X,Y 独立 .

例 4 .11 设 X,Y 是两个随机变量,已知 $E(X)=2$,$E(X^2)=20$,$E(Y)=3$,$E(Y^2)=34$,$\rho_{XY}=0.5$.

求:(1) $E(3X+2Y)$,$E(X-Y)$;

(2) $D(3X+2Y)$,$D(X-Y)$.

解 (1) 因为 $E(X)=2,\quad E(Y)=3$,

所以 $E(3X+2Y)=3E(X)+2E(Y)=12$,

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

(2) 因为 $E(X^2)=20$,$E(Y^2)=34$,

所以 $$D(X)=E(X^2)-[E(X)]^2=20-4=16,$$
$$D(Y)=E(Y^2)-[E(Y)]^2=34-9=25.$$
又因为
$$\operatorname{cov}(X,Y)=\rho_{XY}\quad D(X)\quad D(Y)=0.5\times4\times5=10,$$
$$D(3X+2Y)=D(3X)+D(2Y)+2\operatorname{cov}(3X,2Y)$$
$$=9D(X)+4D(Y)+12\operatorname{cov}(X,Y)=364.$$
$$D(X-Y)=D(X)+D(Y)-2\operatorname{cov}(X,Y)=24.$$

习 题 选 解

(4－1)　已知随机变量 X 的分布函数为
$$F(x)=\begin{cases}0, & x\leq0;\\ \dfrac{x}{4}, & 0<x\leq4;\\ 1, & x>4.\end{cases}$$
求 $E(X)$.

解　因为 $$f(x)=F'(x)=\begin{cases}\dfrac{1}{4}, & 0<x\leq4;\\ 0, & \text{其他}.\end{cases}$$
所以 $$E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\int_0^4\frac{x}{4}\mathrm{d}x=2.$$

(4－2)　某车间生产的圆盘直径服从均匀分布 $U[a,b]$,**求圆盘面积的期望.**

解　用 X **表示圆盘直径,** Y **表示圆盘面积,则有**
$$Y=\frac{\pi}{4}X^2,\qquad f_X(x)=\begin{cases}\dfrac{1}{b-a}, & a\leq x\leq b;\\ 0, & \text{其他}.\end{cases}$$
$$E(Y)=E\quad\frac{\pi}{4}X^2\quad=\int_{-\infty}^{+\infty}\frac{\pi}{4}x^2f_X(x)\mathrm{d}x$$

$$= \frac{\pi}{4}\int_a^b \frac{x^2}{b-a}\mathrm{d}x = \frac{a^2+ab+b^2}{12}.$$

(4-3) **若有 n 把看上去样子相同的钥匙,其中只有 1 把能打开门上的锁,用它们去试开门上的锁.设取到每只钥匙是等可能的,若每只钥匙试开一次后除去,求试开次数 X 的期望.**

解 设 A_i 表示第 i 次打开门,$i=1,2,\cdots,n$.X 的可能取值为 $1,2,\cdots,n$,并可求出各取值的概率.

$$P\{X=1\}=P(A_1)=\frac{1}{n},$$

$$P\{X=2\}=P\{\overline{A_1}A_2\}=P(\overline{A_1})P(A_2|\overline{A_1})$$
$$=\frac{n-1}{n}\cdot\frac{1}{n-1}=\frac{1}{n},$$

……

$$P\{X=n\}=P(\overline{A_1}\,\overline{A_2}\cdots\overline{A_{n-1}}A_n)$$
$$=P(\overline{A_1})P(\overline{A_2}|A_1)\cdots P(A_n|\overline{A_1}\,\overline{A_2}\cdots\overline{A_{n-1}})$$
$$=\frac{n-1}{n}\cdot\frac{n-2}{n-1}\cdot\cdots\cdot\frac{1}{2}\cdot 1=\frac{1}{n}.$$

因此,随机变量 X 的分布列为

X	1	2	3	$\cdots$	n
p	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	$\cdots$	$\frac{1}{n}$

故 $$E(X)=\sum_{i=1}^{n} i\cdot\frac{1}{n}=\frac{n(n+1)}{2}\cdot\frac{1}{n}=\frac{n+1}{2}.$$

(4-4) **设随机变量 X 的分布函数为**

$$F(x)=\begin{cases}1-\dfrac{a^3}{x^3}, & x\geq a;\\ 0, & x<a.\end{cases}$$

其中 $a>0$.求 $E(X)$.

解 **因为** $$f(x)=F'(x)=\begin{cases}\dfrac{3a^3}{x^4}, & x\geq a;\\ 0, & x<a.\end{cases}$$

所以
$$E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x=\int_{a}^{+\infty}\frac{3a^3}{x^3}\mathrm{d}x$$
$$=3a^3\left[-\frac{1}{2x^2}\right]_a^{+\infty}=\frac{3}{2}a$$

(4-5)　设舰艇横向摇摆的随机振幅 X 服从瑞利分布,其概率密度为

$$f(x)=\begin{cases}\dfrac{x}{\sigma^2}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}, & x>0;\\ 0, & x\leq 0.\end{cases}$$

其中 $\sigma>0$.求 $E(X)$.

解　$E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x=\int_{0}^{+\infty} x\dfrac{x}{\sigma^2}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\mathrm{d}x$

$$=-\int_0^{+\infty} x\mathrm{d}\left(\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\right)=-x\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\Big|_0^{+\infty}+\int_0^{+\infty}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\mathrm{d}x$$

$$\overset{\frac{x}{\sigma}=t}{=}\sigma\int_0^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\sqrt{\frac{\pi}{2}}\sigma.$$

(4-6)　设 (X,Y) 的概率密度为

$$f(x,y)=\begin{cases}12y^2, & 0\leq y\leq x\leq 1;\\ 0, & \text{其他}.\end{cases}$$

求:(1) **关于 X,Y 的边缘概率密度** $f_X(x), f_Y(y)$;

(2) $E(X), E(Y)$;

(3) $E(XY)$;

(4) $E(X^2+Y^2)$.

解　(1) $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$

$$=\begin{cases}\int_0^x 12y^2\mathrm{d}y, & 0\leq x\leq 1;\\ 0, & \text{其他}.\end{cases}$$

$$=\begin{cases}4x^3, & 0\leq x\leq 1;\\ 0, & \text{其他}.\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x$$

$$=\begin{cases}\int_y^1 12y^2\mathrm{d}x, & 0\le y\le 1;\\ 0, & \text{其他}.\end{cases}$$

$$=\begin{cases}12y^2(1-y), & 0\le y\le 1;\\ 0, & \text{其他}.\end{cases}$$

(2) $E(X)=\int_{-\infty}^{+\infty} xf_X(x)\mathrm{d}x=\int_0^1 x\cdot 4x^3\mathrm{d}x=\dfrac{4}{5}$;

$E(Y)=\int_{-\infty}^{+\infty} yf_Y(y)\mathrm{d}y=\int_0^1 y\cdot 12y^2(1-y)\mathrm{d}y=\dfrac{3}{5}$.

(3) $E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)\mathrm{d}x\mathrm{d}y$

$$=\int_0^1\mathrm{d}x\int_0^x xy\cdot 12y^2\mathrm{d}y=\frac{1}{2}.$$

(4) $E(X^2+Y^2)=E(X^2)+E(Y^2)$

$$=\int_0^1 x^2\cdot 4x^3\mathrm{d}x+\int_0^1 y^2\cdot 12y^2(1-y)\mathrm{d}y$$

$$=\frac{16}{15}.$$

(4－7)　一民航机场的送客汽车每次载 20 位旅客自机场开出.沿途有 10 个车站,若到达一个车站没有旅客下车,就不停车.假设每位旅客在各个车站下车是等可能的,求每趟汽车的平均停车次数.

解　定义随机变量

$$X_i=\begin{cases}1, & \text{第 } i \text{ 站有人下车};\\ 0, & \text{第 } i \text{ 站无人下车}.\end{cases}\quad i=1,2,\cdots,10.$$

X 表示汽车停车总次数,则有

$$X=\sum_{i=1}^{10}X_i,\quad \text{故 } E(X)=\sum_{i=1}^{10}E(X_i).$$

$$P\{X_i=0\}=P\{20\text{ 个旅客都不在第 } i \text{ 站下车}\}=\frac{9^{20}}{10^{20}},$$

$$P\{X_i=1\}=P\{20\text{ 个旅客在第 }i\text{ 站至少有一人下车}\}$$

$$=1-P\{X_i=0\}=1-\left(\frac{9}{10}\right)^{20}.$$

故 $$E(X_i)=0\times\left(\frac{9}{10}\right)^{20}+1\times\left[1-\left(\frac{9}{10}\right)^{20}\right]=1-\left(\frac{9}{10}\right)^{20},$$

从而 $$E(X)=\sum_{i=1}^{10}E(X_i)=10\left[1-\left(\frac{9}{10}\right)^{20}\right]\approx 8.784.$$

(4-8)　设二维随机变量 (X, Y) 服从圆域 $G: x^2+y^2\leq R^2$ 上的均匀分布，令 $Z=\sqrt{X^2+Y^2}$，求 $E(Z)$.

解 $$E(Z)=\iint\limits_G\sqrt{x^2+y^2}\cdot\frac{1}{\pi R^2}\,\mathrm{d}x\,\mathrm{d}y$$

$$=\frac{1}{\pi R^2}\int_0^{2\pi}\mathrm{d}\theta\int_0^R r^2\,\mathrm{d}r=\frac{2}{3}R.$$

(4-9)　设随机变量 X, Y 相互独立，它们的概率密度分别为

$$f_X(x)=\begin{cases}2\mathrm{e}^{-2x}, & x>0;\\ 0, & x\leq 0.\end{cases}\qquad f_Y(y)=\begin{cases}4\mathrm{e}^{-4y}, & y>0;\\ 0, & y\leq 0.\end{cases}$$

求 $D(X+Y)$.

解　由于 X 与 Y 是相互独立的，所以有

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

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

$$=\int_0^{+\infty}x^2\,2\mathrm{e}^{-2x}\,\mathrm{d}x-\left(\int_0^{+\infty}x\,2\mathrm{e}^{-2x}\,\mathrm{d}x\right)^2=\frac{1}{4}.$$

$$D(Y)=E(Y^2)-[E(Y)]^2$$

$$=\int_0^{+\infty}y^2\,4\mathrm{e}^{-4y}\,\mathrm{d}y-\left(\int_0^{+\infty}y\,4\mathrm{e}^{-4y}\,\mathrm{d}y\right)^2=\frac{1}{16}.$$

故 $$D(X+Y)=D(X)+D(Y)=\frac{1}{4}+\frac{1}{16}=\frac{5}{16}.$$

(4-10)　设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，且服从同一分布，期望为 μ，方差为 σ^2. 令 $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$.

求:(1) $E(\overline{X})$; (2) $D(\overline{X})$.

解 (1) $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$

$$= \frac{1}{n}n\mu = \mu .$$

(2) $D(\overline{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i)$

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} .$$

(4-11) **一台设备由三大部件构成,在设备运转中部件需要调整的概率分别为** 0.1,0.2,0.3.**假设各部件的状态相互独立,以** X **表示同时需要调整的部件数,试求** X **的期望与方差**.

解 **由题设可知**, X **的可能取值为** 0,1,2,3.

$P\{X=0\}=0.9\times0.8\times0.7=0.504$,

$P\{X=1\}=0.1\times0.8\times0.7+0.9\times0.2\times0.7+0.9\times0.8\times0.3$

$=0.398$,

$P\{X=2\}=0.1\times0.2\times0.7+0.1\times0.8\times0.3+0.9\times0.2\times0.3$

$=0.092$.

$P\{X=3\}=0.1\times0.2\times0.3=0.006$,

$E(X)=0\times0.504+1\times0.398+2\times0.092+3\times0.006$

$=0.6$,

$E(X^2)=0^2\times0.504+1^2\times0.398+2^2\times0.092+3^2\times0.006$

$=0.82$,

$D(X)=E(X^2)-[E(X)]^2=0.82-0.6^2=0.46$.

(4-12) **设** X, Y **相互独立,且** $E(X)=E(Y)=1$, $D(X)=2$, $D(Y)=4$,**求** $E[(X+Y)^2]$.

解 **由** X **与** Y **的独立性有**

$$E[(X+Y)^2]=E(X^2+2XY+Y^2)$$

$$
\begin{aligned}
&= E(X^2) + 2E(X)E(Y) + E(Y^2) \\
&= D(X) + [E(X)]^2 + 2E(X)E(Y) \\
&\quad + D(Y) + [E(Y)]^2 \\
&= 2 + 1 + 2 + 4 + 1 = 10 .
\end{aligned}
$$

(4－13)　设二维连续随机向量(X, Y)的概率密度为

$$
f(x, y) = \begin{cases} \dfrac{1}{8}(x + y), & 0 \leq x \leq 2, 0 \leq y \leq 2; \\ 0, & \text{其他} . \end{cases}
$$

求 $E(X), E(Y), \operatorname{cov}(X, Y), \rho_{XY}$.

解　$$E(X) = \int_0^2\int_0^2 \frac{x}{8}(x + y)\,\mathrm{d}x\mathrm{d}y = \frac{7}{6},$$

$$E(Y) = \int_0^2\int_0^2 \frac{y}{8}(x + y)\,\mathrm{d}x\mathrm{d}y = \frac{7}{6} .$$

因为　$$\operatorname{cov}(X, Y) = E(XY) - E(X)E(Y),$$

而　$$E(XY) = \int_0^2\int_0^2 \frac{xy}{8}(x + y)\,\mathrm{d}x\mathrm{d}y = \frac{4}{3},$$

所以　$$\operatorname{cov}(X, Y) = \frac{4}{3} - \left(\frac{7}{6}\right)^2 = -\frac{1}{36} .$$

为求 ρ_{XY},必须先求 $D(X), D(Y)$.

$$E(X^2) = \frac{1}{8}\int_0^2\int_0^2 x^2(x + y)\,\mathrm{d}x\mathrm{d}y = \frac{5}{3},$$

$$D(X) = E(X^2) - [E(X)]^2 = \frac{5}{3} - \left(\frac{7}{6}\right)^2 = \frac{11}{36},$$

同理可求得 $D(Y) = \dfrac{11}{36}$,

故　$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{D(X) \cdot D(Y)} = -\frac{1}{36}\bigg/\frac{11}{36} = -\frac{1}{11} .$$

(4－14)　设随机变量 X 与 Y 相互独立,且 X 与 Y 有相同的概率分布 .记 $U = X + Y$, $V = X - Y$.**证明** $\rho_{UV} = 0$.

证　$\rho_{UV} = \dfrac{\operatorname{cov}(X, Y)}{D(U)\quad D(V)}$,**只需计算分子** $\operatorname{cov}(U, V)$.**而**

$$\operatorname{cov}(U, V) = E(UV) - E(U)E(V)$$

$$= E[(X+Y)(X-Y)] - E(X+Y)E(X-Y)$$

$$= E(X^2) - E(Y^2) - [E(X)+E(Y)] \times [E(X)-E(Y)]$$

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

$$= D(X) - D(Y) = 0 .$$

(因为 X, Y 同分布,故 $D(X) = D(Y)$.)

(4－15) 设随机变量 X 的方差 $D(X) = 16$,随机变量 Y 的方差 $D(Y) = 25$,又 X 与 Y 的相关系数 $\rho_{XY} = 0.5$.

求 $D(X+Y), D(X-Y)$.

解 $D(X+Y) = D(X) + D(Y) + 2\mathrm{cov}(X, Y)$

$$= D(X) + D(Y) + 2\rho_{XY} \quad D(X) \quad D(Y)$$

$$= 16 + 25 + 2\times 0.5\times 4\times 5 = 61 .$$

$$D(X-Y) = D(X) + D(Y) + 2\mathrm{cov}(X, -Y)$$

$$= D(X) + D(Y) - 2\rho_{XY} \quad D(X) \quad D(Y)$$

$$= 16 + 25 - 2\times 0.5\times 4\times 5 = 21 .$$

(4－16) 已知随机变量 X 服从标准正态分布 $N(0,1)$,令 $Y = X^3$,求 X 与 Y 的相关系数 ρ_{XY} .

解 $$D(X) = 1,$$

$$D(Y) = E(Y^2) - [E(Y)]^2 = E(X^6) - [E(X^3)]^2 ,$$

$$E(X^6) = \int_{-\infty}^{+\infty} x^6 f_X(x)\mathrm{d}x = \int_{-\infty}^{+\infty} x^6 \frac{1}{2\pi} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = 5\times 3 = 15^{①} ,$$

又 $E(Y) = E(X^3) = 0$,从而 $D(Y) = 15$.

$$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y) = E(XY)$$

$$= E(X^4) = \int_{-\infty}^{+\infty} x^4 \frac{1}{2\pi} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = 3,$$

---

① 其值是 X 的 6 阶原点矩,见教材第 143 页 .

故 $$\operatorname{cov}(X, Y) = 3,$$

$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{3}{1 \cdot \sqrt{15}} = \frac{3\sqrt{15}}{15} = \frac{1}{5}\sqrt{15}.$$

(4-17) **设 A 和 B 是两个事件**, $P(A) > 0$, $P(B) > 0$. **定义随机变量 X, Y 如下:**

$$X = \begin{cases} 1, & \text{若 } A \text{ 发生;} \\ 0, & \text{若 } A \text{ 不发生.} \end{cases} \qquad Y = \begin{cases} 1, & \text{若 } B \text{ 发生;} \\ 0, & \text{若 } B \text{ 不发生.} \end{cases}$$

证明:若相关系数 $\rho_{XY} = 0$,则 X 与 Y 必定独立.

证 由题设,可写出 X、Y 及 XY 的分布列如下:

X	0	1
p	$P(\overline{A})$	$P(A)$

Y	0	1
p	$P(\overline{B})$	$P(B)$

XY	0	1
p	$1 - P(AB)$	$P(AB)$

由上面的分布列可计算出

$$E(X) = 0 \times P(\overline{A}) + 1 \times P(A) = P(A),$$

$$E(Y) = 0 \times P(\overline{B}) + 1 \times P(B) = P(B),$$

$$E(XY) = 0 \times [1 - P(AB)] + 1 \times P(AB) = P(AB).$$

若相关系数 $\rho_{XY} = 0$,则 $\operatorname{cov}(X, Y) = 0$,即

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

则 $$P(AB) = P(A)P(B).$$

故事件 A 与事件 B 是相互独立的.

因 A, B 独立时, A 与 $\overline{B}$, $\overline{A}$ 与 B 及 $\overline{A}$ 与 $\overline{B}$ 也独立,故可得

$$P\{X = 1, Y = 1\} = P(AB) = P(A)P(B) = P\{X = 1\}P\{Y = 1\}.$$

$$P\{X = 1, Y = 0\} = P(A\overline{B}) = P(A)P(\overline{B}) = P\{X = 1\}P\{Y = 0\}.$$

同理可证 $$P\{X = 0, Y = 1\} = P\{X = 0\}P\{Y = 1\},$$

$$P\{X = 0, Y = 0\} = P\{X = 0\}P\{Y = 0\}.$$

即对所有可能的(x_i, y_j),有

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

故 X 与 Y 相互独立.

(**4－18**) **设随机变量 X 与 Y 相互独立,且都服从正态分布 $N(0, \sigma^2)$,记 $U = \alpha X + \beta Y$, $V = \alpha X - \beta Y$(α, β 是不相等的常数). 求 U 与 V 的相关系数 ρ_{UV}.**

解 $$\rho_{UV} = \frac{\operatorname{cov}(U, V)}{D(U) \quad D(V)} = \frac{E(UV) - E(U)E(V)}{D(U) \quad D(V)},$$

下面分别求 $E(U)$, $E(V)$, $D(U)$, $D(V)$, $E(UV)$.

已知 $$E(X) = E(Y) = \mu, \quad D(X) = D(Y) = \sigma^2.$$

$$E(U) = E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y) = (\alpha + \beta)\mu,$$

$$E(V) = E(\alpha X - \beta Y) = (\alpha - \beta)\mu.$$

因 X 与 Y 相互独立,故

$$D(U) = D(\alpha X + \beta Y) = \alpha^2 D(X) + \beta^2 D(Y) = (\alpha^2 + \beta^2)\sigma^2,$$

$$D(V) = D(\alpha X - \beta Y) = \alpha^2 D(X) + (-\beta)^2 D(Y) = (\alpha^2 + \beta^2)\sigma^2,$$

$$\begin{aligned}
E(UV) &= E[(\alpha X + \beta Y)(\alpha X - \beta Y)] = E[\alpha^2 X^2 - \beta^2 Y^2] \\
&= E(\alpha^2 X^2) - E(\beta^2 Y^2) = \alpha^2 E(X^2) - \beta^2 E(Y^2) \\
&= \alpha^2[D(X) + E^2(X)] - \beta^2[D(Y) + E^2(Y)] \\
&= \alpha^2(\sigma^2 + \mu^2) - \beta^2(\sigma^2 + \mu^2) = (\alpha^2 - \beta^2)(\sigma^2 + \mu^2),
\end{aligned}$$

$$E(U)E(V) = (\alpha + \beta)\mu \cdot (\alpha - \beta)\mu = (\alpha^2 - \beta^2)\mu^2.$$

故 $$\rho_{UV} = \frac{(\alpha^2 - \beta^2)(\sigma^2 + \mu^2) - (\alpha^2 - \beta^2)\mu^2}{\sigma(\alpha^2 + \beta^2)} = \frac{\alpha^2 - \beta^2}{\alpha^2 + \beta^2}.$$

试 题 选 解

[**4－1**] **是非题**

1. **设 ρ 是随机变量 X 与 Y 的相关系数,则 $|\rho| = 1$ 的充要条件是存在常数 a, b,使 $P\{Y = a + bX\} = 1$.**

2．设 X, Y 不相关，则 $E(XY)=E(X)E(Y)$．

3．设 X_1, X_2, X_3 独立同分布均服从 $N(\mu,\sigma^2)$，若 $Y=a_1X_1-a_2X_2-a_3X_3$，a_1, a_2, a_3 为常数，则 $D(Y)=(a_1^2+a_2^2+a_3^2)\sigma^2$．

4．设 $X\sim N(-3,1)$，$Y\sim N(2,1)$，则对 $Z=X-2Y+7$ 有 $X(Z)=0$．

解　1．正确．见教材第 138 页相关系数的性质(2)．

2．正确．由不相关的定义可推得．

3．正确．由方差的性质可得．

4．正确．由数学期望的性质可得．

[**4－2**]　填空题

1．设 $D(X)=1$，$D(Y)=6$，$\mathrm{cov}(X,Y)=2$，则 $D(X-Y)=$__________．

2．设 X 与 Y 独立，则它们的相关系数 $\rho(X,Y)=$__________．

3．设 $E(X)=a$，$E\left(\frac{1}{Y}\right)=b$，且 X 与 Y 相互独立，则 $E\left(\frac{X}{Y}\right)=$__________．

4．设 X 服从 $n=10$，$p=0.4$ 的二项分布，则 $E(X^2)=$__________．

5．已知随机变量 X 和 Y 相互独立，且它们分别在区间$[-1,3]$和$[2,4]$上服从均匀分布，则 $E(XY)=$__________．

解　1．3．因为 $D(X-Y)=D(X)+D(Y)-2\mathrm{cov}(X,Y)$．

2．0．因为独立一定不相关．

3．ab．因为 X 与 Y 独立，所以 X 与 $\frac{1}{Y}$ 也独立．

4．18.4．由于二项分布的期望和方差已知，而 $E(X^2)=D(X)+[E(X)]^2$．

5．3．因为 $E(XY)=E(X)E(Y)$．

[**4－3**]　在贝努利试验中，试验成功的概率为 p，试验进行到成功与失败都出现为止，求试验次数的数学期望．

解 设 X 为成功与失败都出现的试验次数,则

$$P\{X=n\}=p^{n-1}q+pq^{n-1},\quad n\geq 2.$$

所以,
$$\begin{aligned}E(X)&=\sum_{n=2}^{\infty}n(p^{n-1}q+pq^{n-1})\\&=q\cdot\frac{\mathrm{d}}{\mathrm{d}p}\left[\sum_{n=2}^{\infty}p^n\right]+p\cdot\frac{\mathrm{d}}{\mathrm{d}q}\left[\sum_{n=2}^{\infty}q^n\right]\\&=q\cdot\frac{\mathrm{d}}{\mathrm{d}p}\left[\frac{p^2}{1-p}\right]+p\cdot\frac{\mathrm{d}}{\mathrm{d}q}\left[\frac{q^2}{1-q}\right]\\&=\frac{p^2+q^2+pq}{pq}.\end{aligned}$$

其中
$$q=1-p.$$

[4-4] **设随机变量 X 在 $[0,1]$ 上均匀分布, Y 在 $[1,3]$ 上均匀分布,且 X 与 Y 独立,求 $E(XY)$ 与 $D(XY)$.**

解 由于 $E(X)=\frac{1}{2}$, $E(Y)=2$ 及独立性有 $D(X)=\frac{1}{12}$, $D(Y)=\frac{1}{3}$.

$$E(XY)=E(X)E(Y)=\frac{1}{2}\times 2=1,$$

$$\begin{aligned}D(XY)&=E(XY-1)^2=E(X^2Y^2)-2E(XY)+1\\&=E(X^2)E(Y^2)-2+1\\&=[D(X)+(E(X))^2][D(Y)+(E(Y))^2]-1\\&=\frac{1}{3}\cdot\frac{13}{3}-1=\frac{4}{9}.\end{aligned}$$

[4-5] **一工厂生产的某种设备的寿命 X(以年计)服从指数分布,密度函数为**

$$f(x)=\begin{cases}\frac{1}{4}\mathrm{e}^{-\frac{x}{4}}, & x>0;\\ 0, & x\leq 0.\end{cases}$$

为确保消费者的利益,工厂规定出售的设备若在一年内损坏可以调换.若售出一台设备,工厂获利 100 元,而调换一台则损失 200 元.试求工厂出售一台设备赢利的数学期望.

解 由于

$$P\{X\geq 1\}=\int_1^{+\infty}\frac{1}{4}e^{-\frac{x}{4}}dx=e^{-\frac{1}{4}}.$$

售出一台设备的赢利情况为

$$Y=\begin{cases}100, & x\geq 1;\\ -200, & x<1.\end{cases}$$

因此出售一台设备赢利的数学期望为

$$\begin{aligned}E(Y)&=100\times P\{X\geq 1\}-200\times P\{X<1\}\\&=300e^{-\frac{1}{4}}-200(1-e^{-\frac{1}{4}})\approx 33.64.\end{aligned}$$

[4－6] 设 X, Y 相互独立,其概率密度为

$$f_X(x)=\begin{cases}2x, & 0\leq x\leq 1;\\ 0, & \text{其他}.\end{cases}\qquad f_Y(y)=\begin{cases}e^{-(y-5)}, & y>5;\\ 0, & \text{其他}.\end{cases}$$

求 $E(XY)$和 $D(X+Y)$.

解 (1) 因为 X 与 Y 独立,所以 $E(XY)=E(X)E(Y)$

$$E(X)=\int_0^1 2x^2dx=\frac{2}{3},$$

$$E(Y)=\int_5^{+\infty}ye^{-(y-5)}dy=e^5\int_5^{+\infty}ye^{-y}dy=6.$$

故 $$E(XY)=E(X)E(Y)=\frac{2}{3}\times 6=4.$$

(2) 由 X 与 Y 的独立性有 $D(X+Y)=D(X)+D(Y)$,

而 $D(X)=E(X^2)-[E(X)]^2$, $D(Y)=E(Y^2)-[E(Y)]^2$,

由 $E(X^2)=\int_0^1 2x^3dx=\frac{1}{2}$,得 $D(X)=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}$.

$$E(Y^2)=\int_5^{+\infty}y^2e^{-(y-5)}dy=e^5\int_5^{+\infty}y^2e^{-y}dy=37,$$

得 $D(Y)=37-6^2=1$,故 $D(X+Y)=\frac{1}{18}+1=\frac{19}{18}$.

[4－7] 一台设备由三大部件构成,在设备运转中各部件需调整的概率相应为 0.10, 0.20 和 0.30.假设各部件状态相互独

立，以 X 表示同时需调整的部件数.

(1) 试求 X 的概率分布；

(2) 求 $E(X)$和 $D(X)$.

解 (1) 设 A_i 表示第 i 个部件需调整，$i=1,2,3$.有 $P(A_1)=0.10$，$P(A_2)=0.20$，$P(A_3)=0.30$.X 所有可能取值为 $0,1,2,3$.

由 A_1,A_2,A_3 独立，有

$$P\{X=0\}=P(\overline{A_1}\,\overline{A_2}\,\overline{A_3})=0.9\times0.8\times0.7=0.504.$$

$$\begin{aligned}P\{X=1\}&=P(A_1\overline{A_2}\,\overline{A_3})+P(\overline{A_1}A_2\overline{A_3})+P(\overline{A_1}\,\overline{A_2}A_3)\\&=0.1\times0.8\times0.7+0.9\times0.2\times0.7\\&\quad+0.9\times0.8\times0.3\\&=0.398.\end{aligned}$$

$$\begin{aligned}P\{X=2\}&=P(A_1A_2\overline{A_3})+P(A_1\overline{A_2}A_3)+P(\overline{A_1}A_2A_3)\\&=0.1\times0.2\times0.7+0.9\times0.2\times0.3\\&\quad+0.1\times0.8\times0.3\\&=0.092.\end{aligned}$$

$$P\{X=3\}=P(A_1A_2A_3)=0.1\times0.2\times0.3=0.006.$$

(2) $$\begin{aligned}E(X)&=0\times0.504+1\times0.398+2\times0.092+3\times0.006\\&=0.6.\end{aligned}$$

$$\begin{aligned}D(X)&=E(X^2)-[E(X)]^2\\&=(1^2\times0.398+2^2\times0.092+3^2\times0.006)-(0.6^2)\\&=0.82-0.36=0.46.\end{aligned}$$

[**4－8**] 设随机变量 X_1 与 X_2 相互独立，且 $X_1\sim N(\mu,\sigma^2)$，$X_2\sim N(\mu,\sigma^2)$.令 $X=X_1+X_2$，$Y=X_1-X_2$.

求：(1) $D(X)$，$D(Y)$；

(2) X 与 Y 的相关系数 ρ_{XY}.

解 $$D(X)=D(X_1+X_2)=D(X_1)+D(X_2)=2\sigma^2,$$

$$D(Y)=D(X_1-X_2)=D(X_1)+D(X_2)=2\sigma^2,$$

$$\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)$$

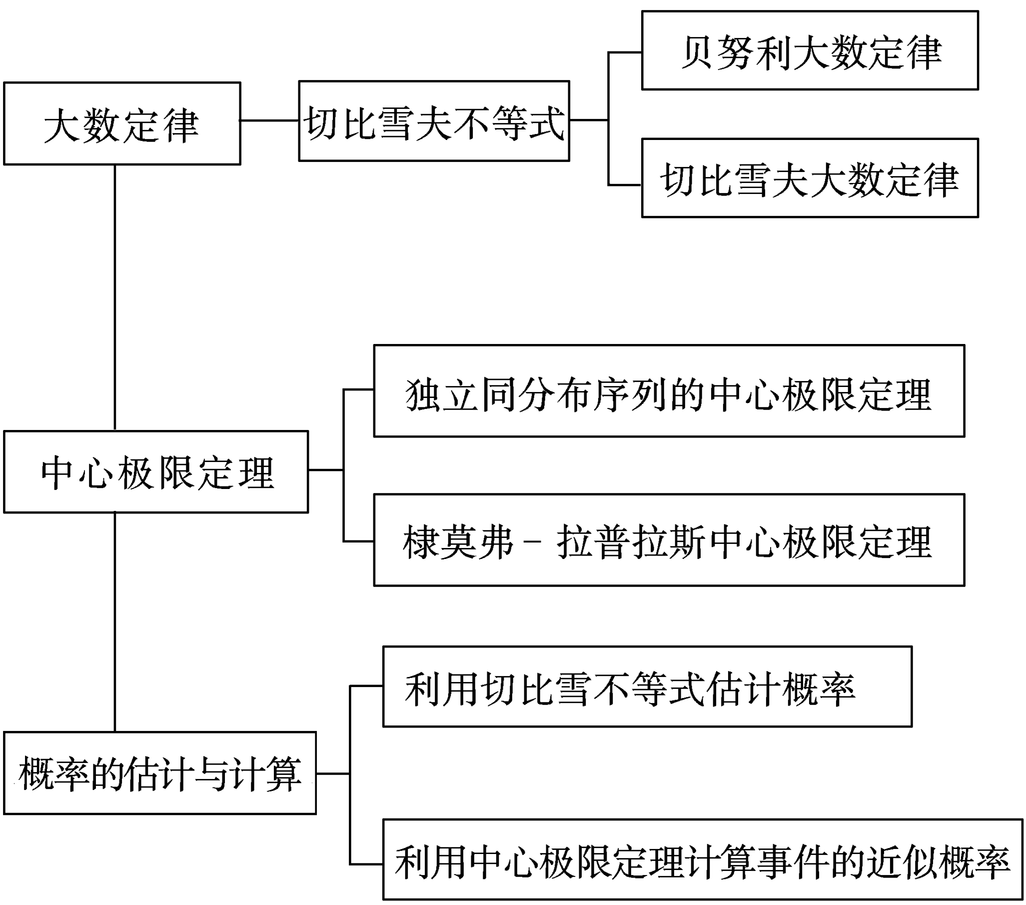
$$
\begin{aligned}
&= E[(X_1 + X_2)(X_1 - X_2)] - E(X_1 + X_2) \\
&\quad \cdot E(X_1 - X_2) \\
&= E(X_1^2) - E(X_2^2) - [E(X_1) + E(X_2)] \\
&\quad \cdot [E(X_1) - E(X_2)] \\
&= D(X_1) - D(X_2) = 0 .
\end{aligned}
$$

则
$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{D(X) \quad D(Y)} = 0 .$$

第五章　大数定律与中心极限定理

内 容 提 要

一、知识脉络



二、知识要点

(一) 大数定律

1. 切比雪夫不等式

设随机变量 X 有期望 $E(X)$ 与方差 $D(X)$,则对任意正数 ε,有

$$P\{X - E(X) \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2},$$

或

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

2. 贝努利大数定律

设 m 是 n 次独立重复试验中 A 发生的次数, p 是事件 A 的概率 .$P(A) = p$,则对任意正数 ε,有

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

3. 切比雪夫大数定律

设独立随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$,服从相同的分布, $E(X_i) = \mu, D(X_i) = \sigma^2 (i = 1, 2, \cdots)$.则对任意正数 ε,有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right\} = 1.$$

(二) 中心极限定理

1. 独立同分布序列的中心极限定理

设相互独立的随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$服从相同的概率分布,且 $E(X_i) = \mu, D(X_i) = \sigma^2 (i = 1, 2, \cdots)$,记

$$Z_n^0 = \frac{\sum_{i=1}^{n} X_i - n\mu}{n\sigma},$$

设 Z_n^0 的分布函数为

$$F_n(x) = P\{Z_n^0 \leq x\},$$

则有 $$\lim_{n\to\infty} F_n(x) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = \Phi(x).$$

2 . 棣莫弗-拉普拉斯中心极限定理

设 m 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 发生的概率,则对任意 x,有

$$\lim_{n\to\infty} P\ \frac{m-np}{npq}\leq x\ = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = \Phi(x),$$

其中 $q=1-p$.

要 点 范 例

一、利用切比雪夫不等式估计概率

例 5 .1 设正常男性成人每毫升血液中,白细胞数为 X,且已知 $E(X)=7\ 300$, $\sigma(X)=700$.利用切比雪夫不等式估计每毫升血液含白细胞数在5 200至9 400之间的概率 .

解 设正常男性成人每毫升血液中含白细胞数为 X,由题设

$$E(X)=7\ 300,\quad D(X)=700^2 .$$

由切比雪夫不等式

$$\begin{aligned}
&P\{5\ 200<X<9\ 400\}\\
&=P\{-2\ 100<X-7\ 300<2\ 100\}\\
&=P\{|X-7\ 300|<2\ 100\}\\
&\geq 1-\frac{700^2}{2\ 100^2}=\frac{8}{9}.
\end{aligned}$$

二、利用中心极限定理计算事件的近似概率

例 5 .2 计算机进行加法计算时,把每个加数取为最接近于它的整数来计算 .设所有取整误差是相互独立的随机变量,且服从

均匀分布 $U[-0.5, 0.5]$.求 300 个数相加时误差总和的绝对值小于 10 的概率 .

解　设 X_i 为第 i 个加数的取整误差,则 $X_i \sim U[-0.5, 0.5]$.

$$E(X_i) = 0, \quad D(X_i) = \frac{1}{12} \quad (i = 1, 2, \cdots, 300) .$$

有

$$P \left| \sum_{i=1}^{300} X_i \right| < 10 = P \frac{\left| \sum_{i=1}^{300} X_i \right|}{300 \times \frac{1}{12}} < 2$$

$$\approx \Phi(2) - \Phi(-2) = 0.954\,4 .$$

注　①对于独立同分布随机变量 X_i,在求得 $E(X_i)$ 和 $D(X_i)$条件下,可用独立同分布序列的中心极限定理计算概率的近似值 .

② 由中心极限定理可知,当 n 很大时,有

$$P \quad a < \frac{\sum_{i=1}^{n} X_i - nE(X_i)}{nD(X)} \leq b \approx \Phi(b) - \Phi(a) .$$

计算概率近似值可直接用这个结论 .

例 5.3　100 台车床彼此独立地工作着,每台车床的实际工作时间占全部工作时间的 80%,求任一时刻有 70 台至 86 台车床在工作的概率 .

解　车床工作的概率为 0.8,车床台数为 $n = 100$.设实际工作台数为 m,由棣莫弗-拉普拉斯中心极限定理,所求概率为

$$P\{70 \leq m \leq 86\}$$

$$\approx \Phi \frac{86 - 100 \times 0.8}{100 \times 0.8 \times 0.2} - \Phi \frac{70 - 100 \times 0.8}{100 \times 0.8 \times 0.2}$$

$$= \Phi(1.5) - \Phi(-2.5) = \Phi(1.5) - (1 - \Phi(2.5))$$

$$= 0.933\,2 - (1 - 0.993\,8) = 0.927 .$$

习 题 选 解

(5－1)　设随机变量 X **的期望** $E(X)=\mu$**,方差** $D(X)=\sigma^2$**,试用切比雪夫不等式估计**

$$P\{|X-\mu|<3\sigma\}.$$

解　取 $\varepsilon=3\sigma$**,则**

$$P\{|X-\mu|<3\sigma\}\geq 1-\frac{\sigma^2}{3^2\sigma^2}=1-\frac{1}{9}=\frac{8}{9}.$$

(5－2)　在每次试验中,事件 A **发生的概率等于** 0.5**.利用切比雪夫不等式估计:在** 1 000 **次试验中,事件** A **发生次数在** 400 **次至** 600 **次之间的概率**.

解　$n=1\ 000$,　$p=0.5$,　$P\left\{\left|\frac{m}{n}-p\right|<\varepsilon\right\}\geq 1-\frac{p(1-p)}{n\varepsilon^2}$,

所以

$$\begin{aligned}P\{400<m<600\}&=P\left\{0.4<\frac{m}{1\ 000}<0.6\right\}\\&=P\left\{\left|\frac{m}{1\ 000}-0.5\right|<0.1\right\}\geq 1-\frac{1}{4\times 1\ 000\times 0.01}\\&=1-\frac{1}{40}=\frac{39}{40}=0.975.\end{aligned}$$

(5－3)　设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ **相互独立,服从相同的分布,** $E(X_i)=0$, $D(X_i)=\sigma^2$**,且** $E(X_i^4)$ **存在**$(i=1,2,\cdots)$**.试证明:对任意** $\varepsilon>0$**,有**

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i^2-\sigma^2\right|<\varepsilon\right\}=1.$$

证　令 $Y_i=X_i^2(i=1,2,\cdots)$**,则**$\{Y_i\}$**也为独立同分布序列**.

$$E(Y_i)=E(X_i^2)=D(X_i)+(E(X_i))^2=\sigma^2+0=\sigma^2,$$

$$D(Y_i)=D(X_i^2)=E(X_i^4)-(E(X_i^2))^2=E(X_i^4)-\sigma^4<+\infty.$$

由独立同分布序列的切比雪夫大数定律,对任意正数 ε,有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i^2-\sigma^2\right|<\varepsilon\right\}=1.$$

(5-4)　设 $X_i(i=1,2,\cdots,50)$是相互独立的随机变量,且都服从泊松分布 $P(0.03)$.令 $\overline{Z}=\sum_{i=1}^{50}X_i$. 试用中心极限定理计算 $P\{\overline{Z}\geq 3\}$.

解　因 $X_i\sim P(0.03)$,**则**

$$E(X_i)=0.03,\quad D(X_i)=0.03,$$

$$P\{\overline{Z}\geq 3\}=1-p\{\overline{Z}<3\}\approx 1-\Phi\left(\frac{3-0.03\times 50}{\sqrt{0.03\times 50}}\right)$$

$$=1-\Phi(1.225)=0.1103.$$

(5-5)　已知一本 300 **页的书中每页印刷错误的个数服从泊松分布** $P(0.2)$.**求这本书的印刷错误总数不多于** 70 **的概率**.

解　设第 i 页的印刷错误个数为 $X_i(i=1,2,\cdots,300)$,**则** $E(X_i)=0.2$, $D(X_i)=0.2$,**且** X_i **相互独立,利用中心极限定理所求概率为**

$$P\left\{\sum_{i=1}^{300}X_i\leq 70\right\}\approx\Phi\left(\frac{70-60}{\sqrt{300\times 0.2}}\right)$$

$$=\Phi\left(\frac{\sqrt{15}}{3}\right)=\Phi(1.29)=0.9015.$$

(5-6)　一部件包括 10 **部分,每部分的长度是一个随机变量,相互独立且服从同一分布,其期望是** 2 mm,**标准差是** 0.05 mm.**规定总长度为** 20 ± 0.1 mm **时产品合格,求产品合格的概率**.

解　设第 i 部分长度为 $X_i(i=1,2,\cdots,10)$, $X_1,X_2,\cdots,X_{10}$ **相互独立且服从同一分布**

$$E(X_i)=2,\quad D(X_i)=(0.05)^2.$$

故由中心极限定理,产品合格的概率为

$$P\left\{20-0.1<\sum_{i=1}^{10}X_i<20+0.1\right\}$$

$$\approx\Phi\left(\frac{0.1}{0.05\sqrt{10}}\right)-\Phi\left(\frac{-0.1}{0.05\sqrt{10}}\right)$$

$$=\Phi\left(\frac{\sqrt{10}}{5}\right)-\Phi\left(-\frac{\sqrt{10}}{5}\right)=2\Phi\left(\frac{\sqrt{10}}{5}\right)-1$$

$$=2\Phi(0.63)-1=0.4714.$$

(**5-7**) **一个复杂系统由** n **个相互独立起作用的部件所组成.每个部件的可靠性为** 0.9, **且必须至少有** 80% **的部件工作才能使整个系统工作,问** n **至少为多少才能使系统的可靠性为** 95%?

解　设 m **为运行期间** n **个部件保持完好的个数.依题意,求当** $m\geq 0.8n$ **时的概率,要求它不小于** 95%,**即**

$$P\{m\geq 0.8n\}\geq 0.95,$$

有

$$P\{m\geq 0.8n\}\approx 1-\Phi\left(\frac{0.8n-0.9n}{\sqrt{n\times 0.9\times 0.1}}\right)$$

$$=1-\Phi\left(-\frac{\sqrt{n}}{3}\right)\geq 0.95,$$

即 $\Phi\left(\frac{\sqrt{n}}{3}\right)\geq 0.95$, **故** $\frac{\sqrt{n}}{3}\geq 1.645$, **得** $n>25$.**因此** n **至少为** 25 **时才能满足要求**.

(**5-8**) **设** $X_1,X_2,\cdots,X_n\cdots$ **独立同分布**, $E(X_i)=\mu$, $D(X_i)=\sigma^2$, $i=1,2,\cdots$.**证明**:

$$P\{|X-\mu|<\varepsilon\}\approx 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)-1,$$

其中 $X=\frac{1}{n}\sum_{i=1}^{n}X_i$, $\Phi(x)$ **是标准正态分布函数**.

证　$P\{|X-\mu|<\varepsilon\}$

$$=P\{\mu-\varepsilon<\overline{X}<\mu+\varepsilon\}\approx\Phi\left(\frac{\mu+\varepsilon-\mu}{\sigma/\sqrt{n}}\right)-\Phi$$

$$\frac{\mu-\varepsilon-\mu}{\sigma/ n}$$

$$=\Phi\left(\frac{n\varepsilon}{\sigma}\right)-\left[1-\Phi\left(\frac{n\varepsilon}{\sigma}\right)\right]$$

$$=2\Phi\left(\frac{n\varepsilon}{\sigma}\right)-1\ .$$

试 题 选 解

[5 − 1]　填空题

1．设 n_A 为 n 次重复独立试验中事件 A 发生的次数，p 为事件 A 在每次试验中发生的概率，则由大数定律，对于任意的 $\varepsilon>0$，有__________．

答　$\lim\limits_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|>\varepsilon\right\}=0$ ．

因为由贝努利大数定律，对任意正数 ε，有 $\lim\limits_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|\leq\varepsilon\right\}=1$，故

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|>\varepsilon\right\}=1-\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|\leq\varepsilon\right\}=1-1=0\ .$$

2．设随机变量列 $X_1, X_2,\cdots$相互独立，具有相同的数学期望与方差：$E(X_k)=\mu$，$D(X_k)=\sigma^2(k=1,2,\cdots)$，令 $Y_n=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$，则由大数定律，对任意的 $\varepsilon>0$，有__________．

答　$\lim\limits_{n\to\infty} P\{|Y_n-\mu|<\varepsilon\}=1$ ．

因为此处由独立同分布序列的切比雪夫大数定律立即可得．

3．设随机变量序列 $X_1, X_2,\cdots$独立同分布，且 $E(X_1)=a$，

$D(X_1)=\sigma^2$,则 $\lim\limits_{n\to\infty}\left(\dfrac{\sum\limits_{i=1}^{n}X_k-na}{n\sigma}\leq 0\right)=$ ________ .

答 $\dfrac{1}{2}$.

此处由独立同分布序列的中心极限定理得

$$\lim_{n\to\infty}P\left(\frac{\sum\limits_{k=1}^{n}X_k-na}{n\sigma}\leq 0\right)=\Phi(0)=\frac{1}{2}.$$

4 . 设 $X_n\ (n=1,2,\cdots)$ 是独立同分布随机变量序列,且 $E(X_n)=\alpha<\infty$,则由大数定律,对任意的 $\varepsilon>0$,有________ .

答 $\lim\limits_{n\to\infty}P\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i-a\right|<\varepsilon\right)=1$.

同 2 题类似 .

5 . 设随机变量 $X\sim U[0,1]$,由切比雪夫不等式可得 $P\left(\left|X-\dfrac{1}{2}\right|\geq\dfrac{1}{3}\right)\leq$________ .

答 $\dfrac{1}{4}$.

因为 $X\sim U[0,1]$, $E(X)=\dfrac{0+1}{2}=\dfrac{1}{2}$, $D(X)=\dfrac{1}{12}$.由切比雪夫不等式 $P\{|X-E(X)|\geq\varepsilon\}\leq\dfrac{D(X)}{\varepsilon^2}$,

$$P\left(\left|X-\frac{1}{2}\right|\geq\frac{1}{3}\right)\leq\frac{\frac{1}{12}}{\left(\frac{1}{3}\right)^2}=\frac{1}{12}\times 3=\frac{1}{4}.$$

[5 - 2] 单选题

设 $\Phi(x)$为标准正态分布函数,

$$X_i=\begin{cases}1, & \text{事件 } A \text{ 发生;}\\ 0, & \text{事件 } A \text{ 不发生.}\end{cases}\quad i=1,2,\cdots,100.$$

且 $P(A)=0.8$, $X_1, X_2, \cdots, X_{100}$相互独立．令 $Y=\sum_{i=1}^{100} X_i$，则由中心极限定理知 Y 的分布函数 $F(y)$近似于(　　)

① $\Phi(y)$　　　② $\Phi\left(\frac{y-80}{4}\right)$

③ $\Phi(16y+80)$　　　④ $\Phi(4y+80)$

答　选择②．

由中心极限定理知 $Y=\sum_{i=1}^{100} X_i$ 的分布近似于正态分布 $N(n\mu, n\sigma^2)$，即 Y 的分布函数 $F(y)$近似于 $\Phi\left(\frac{y-n\mu}{\sqrt{n}\sigma}\right)$，其中

$$\mu=E(X_i)=0.8,$$

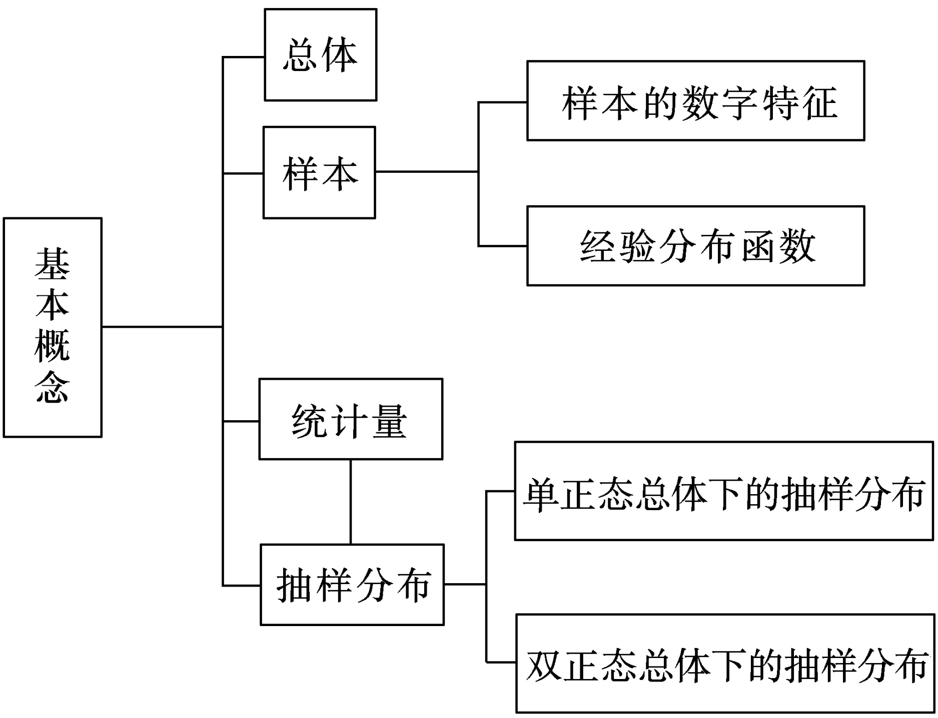
$$D(X)=P(1-P)=0.16,\quad n=100,\quad \sigma=\sqrt{D(X)}=0.4.$$

所以 $F(y)$近似于 $\Phi\left(\frac{y-80}{4}\right)$．

第六章　样本与抽样分布

内 容 提 要

一、知识脉络



二、知识要点

(一) 基本概念

1. 总体

在数理统计中,通常把所研究对象的全体称为总体,总体中的

每个单元称为个体.一般地,总体都联系着一个随机变量 X.

2.样本

在总体 X 中随机抽取的 n 个个体 $X_1, X_2, \cdots, X_n$,称为来自总体 X 的一个样本,n 为样本容量.

若 $X_1, X_2, \cdots, X_n$ 相互独立且与总体 X 有相同的分布,则称 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一个简单随机样本(简称样本).

$X_1, X_2, \cdots, X_n$ 所取的具体数值 $x_1, x_2, \cdots, x_n$ 称为总体 X 的一个样本值[①].

(二) 样本数字特征

设 $x_1, x_2, \cdots, x_n$ 是总体的样本值,则有

1.样本均值

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

2.样本方差

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

3.样本 k 阶原点矩

$$V_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k .$$

4.样本 k 阶中心矩

$$U_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k .$$

(三) 经验分布函数

设总体 X 的分布函数为 $F(x)$,通常是未知的.$x_1, x_2, \cdots, x_n$ 是一容量为 n 的样本值,记样本值中小于或等于 x 的个数为 $m(x)$,定义

① 教材中,样本与样本值均用 $x_1, x_2, \cdots, x_n$ 表示.本书沿用两种形式表示:大写 $X_1, X_2, \cdots, X_n$ 表示样本(随机变量),小写 $x_1, x_2, \cdots, x_n$ 表示样本值.

$$F_n(x) = \frac{m(x)}{n}$$

为 X 的经验分布函数 .

(四) 统计量

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, $g(x_1, x_2, \cdots, x_n)$ 为一个 n 元函数,若 $g(x_1, x_2, \cdots, x_n)$中不含任何未知参数,则称 $g(X_1, X_2, \cdots, X_n)$为统计量 .

$g(X_1, X_2, \cdots, X_n)$的概率分布称为抽样分布 .

(五) 单正态总体的抽样分布

设 $X_1, X_2, \cdots, X_n$ 为来自总体 $X \sim N(\mu, \sigma^2)$的样本,则

(1) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N \quad \mu, \frac{\sigma^2}{n} \quad ;$

(2) $U = \frac{\overline{X} - \mu}{\sigma} \quad n \sim N(0, 1)$;

(3) $\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$;

(4) $\overline{X}$ 与 S^2 相互独立;

(5) $T = \frac{\overline{X} - \mu}{S} \quad n \sim t(n-1)$.

(六) 双正态总体的抽样分布

设 $X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_n$ 分别为来自 $X \sim N(\mu_1, \sigma_1^2)$ 和 $Y \sim N(\mu_2, \sigma_2^2)$的样本,且相互独立 .则

(1) 当 $\sigma_1^2 \neq \sigma_2^2$ 时,

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2} \quad \frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2} \sim t(n_1 + n_2 - 2);$$

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F(n_1 - 1, n_2 - 1);$$

(2) 当 $\sigma_1^2 = \sigma_2^2$ 时,

$$F=\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1).$$

要 点 范 例

一、求样本平均值与样本方差

例 6.1　设样本值如下:

19.1,20.0,21.2,18.8,19.6,20.5,22.0,21.6,19.4,20.3.

求样本平均值 $\overline{x}$、样本方差 S^2、样本二阶中心矩 U_2.

解　为了便于计算,可先估一个大概的平均值 x_0,这里取 x_0 = 20.

$$\begin{aligned}\overline{x}&=20+\frac{1}{10}(-0.9+0+1.2-1.2-0.4\\&\quad+0.5+2+1.6-0.6+0.3)\\&=20.25.\end{aligned}$$

$$\begin{aligned}S^2&=\frac{1}{9}(1.15^2+0.25^2+0.95^2+1.45^2+0.65^2\\&\quad+0.25^2+1.75^2+1.35^2+0.85^2+0.05^2)\\&=1.165.\end{aligned}$$

$$U_2=\frac{1}{10}\times10.485=1.0485.$$

例 6.2　下表是 100 个学生身高测量情况.

身高/cm	154～158	158～162	162～166	166～170	170～174	174～178	178～182
学生数	10	14	26	28	12	8	2

若各组以组中值作为样本中的数值,近似计算样本均值、样本方

差.

解 列表计算如下.

组 别	组中值 x_i	学生数 f_i	分组和 $x_i f_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2 f_i$
154～158	156	10	1 560	-10.1	1 020.1
158～162	160	14	2 240	-6.1	520.94
162～166	164	26	4 264	-2.1	114.66
166～170	168	28	4 704	1.9	101.08
170～174	172	12	2 064	5.9	417.72
174～178	176	8	1 418	9.9	784.08
178～182	180	2	360	13.9	386.42
$\sum$	——	100	16 610	——	3 345.00

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{16\ 610}{100} = 166.1,$$

$$S^2 = \frac{3\ 345}{99} = 33.788\ .$$

注 对于这种分组数据,采用列表计算的方法,简单、清晰且便于检查.

二、正态总体的抽样分布

例 6.3 设总体 $X \sim N(\mu, \sigma^2)$.样本容量 $n = 16$,样本均值 $\overline{x} = 12.5$,样本方差 $S^2 = 5.333\ 3$.

(1) 若已知 $\sigma = 2$,求 $P\{|\overline{x} - \mu| < 0.5\}$;

(2) 若未知 σ,求 $P\{|\overline{x} - \mu| < 0.5\}$.

解 (1) 因为
$$U = \frac{\overline{x} - \mu}{2/\sqrt{16}} \sim N(0,1),$$

所以
$$\begin{aligned} &P\{|\overline{x} - \mu| < 0.5\} \\ &= P\left\{\frac{|\overline{x} - \mu|}{2/\sqrt{16}} < \frac{0.5}{2/\sqrt{16}}\right\} = P\{|U| < 1\} \end{aligned}$$

$$= P\{-1 < U < 1\} = \Phi(1) - \Phi(-1)$$
$$= 2\Phi(1) - 1 = 2\times 0.8413 - 1 \text{(查标准正态表)}$$
$$= 0.682\,6 .$$

(2) 因为 $$t = \frac{\bar{x} - \mu}{S/\sqrt{16}} \sim t(16-1),$$

所以 $$P\{|\bar{x} - \mu| < 0.5\} = P\left\{\frac{|\bar{x} - \mu|}{5.333\,3/\sqrt{16}} < \frac{0.5\times 4}{2.309\,4}\right\}$$
$$= P\{|t| < 0.866\,0\} = 1 - P\{|t| > 0.866\}$$
$$= 1 - 2P\{t > 0.866\,0\} = 1 - 2\times 0.2 \text{(查 } t \text{ 分布表, } n = 15)$$
$$= 0.6 .$$ [①]

例 6.4 设总体 $X\sim N(\mu,\sigma^2)$,已知样本容量 $n = 24$,样本方差 $S^2 = 12.522\,7$,求总体标准差 $\sigma > 3$ 的概率.

解 因为 $\frac{x_i - \bar{x}}{\sigma} \sim N(0,1)$,则有

$$\frac{\sum (x_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

所以 $$P\{\sigma > 3\} = P\left\{\frac{(24-1)S^2}{\sigma^2} < \frac{(24-1)S^2}{9}\right\}$$
$$= P\left\{\chi^2 < \frac{23\times 12.522\,7}{9}\right\} = P\{\chi^2 < 32\}.$$

查 χ^2 分布表得

$$P\{\chi^2 > 32\} = 0.1,$$

所以 $$P\{\sigma > 3\} = 1 - P\{\chi^2 > 32\} = 0.9 .$$

例 6.5 总体 $X\sim N(25,3)$,从中随机抽取两组相互独立样本,样本均值分别为 $\bar{x}$、$\bar{y}$,样本容量分别为 20、30.则 $P\{|\bar{x} - \bar{y}| >$

---

① 根据上侧 α 分位数(以后简称临界值)0.866 和自由度 15 无法从教材后面的 t 分布表查到相应的概率 α,只能从更详细的 t 分布表中查得 α = 0.2.在以后的统计推断中,大家主要是掌握根据给定的概率 α 和自由度 n 查 t 分布表得上侧 α 分位数的方法.

$0.6\} = (\quad)$.

分析 因为 $\bar{x} \sim N\left(25, \frac{3}{20}\right)$, $\bar{y} \sim N\left(25, \frac{3}{30}\right)$, 且 $\bar{x}$、$\bar{y}$ 相互独立, 则有 $D(\bar{x} - \bar{y}) = D(\bar{x}) + D(\bar{y}) = \frac{3}{20} + \frac{3}{30} = \frac{1}{4}$, 所以 $\bar{x} - \bar{y} \sim N\left(0, \frac{1}{4}\right)$.

解

$$\begin{aligned} P\{|\bar{x} - \bar{y}| > 0.6\} &= 1 - P\{|\bar{x} - \bar{y}| \leq 0.6\} \\ &= 1 - P\{-0.6 \leq \bar{x} - \bar{y} \leq 0.6\} \\ &= 1 - \left[\Phi\left(\frac{0.6 - 0}{\sqrt{1/4}}\right) - \Phi\left(\frac{-0.6 - 0}{\sqrt{1/4}}\right)\right] \\ &= 2 - 2 \times \Phi(1.2) = 0.230\,2. \end{aligned}$$

例 6.6 总体 X 服从泊松分布 $P(\lambda)$, 从中抽取样本 $X_1, X_2, \cdots, X_n$. $\bar{X}$ 与 S^2 分别为样本均值与样本方差, 则

(1) $D(\bar{X}) = (\quad)$;

(2) $E(S^2) = (\quad)$.

分析 因为 $X \sim P(\lambda)$, 所以

$$E(X) = \lambda, \quad D(X) = \lambda,$$

$$D(\bar{X}) = \frac{1}{n} D(X) = \frac{\lambda}{n}.$$

又 S^2 为 $D(X) = \lambda$ 的无偏估计, 因此, $E(S^2) = D(X) = \lambda$.

例 6.7 总体 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \cdots, X_n$ 为来自 X 的样本.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad U_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

则 (1) $\dfrac{U_2}{\sigma^2/n} \sim (\quad)$;

(2) $\dfrac{u_2}{(n-1)(\bar{X} - \mu)^2} \sim (\quad)$.

解 (1) $$\frac{U_2}{\sigma^2/n}=\frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1),$$

即
$$\frac{U_2}{\sigma^2/n}\sim\chi^2(n-1).$$

(2) **令** $Y_1=\dfrac{U_2}{\sigma^2/n}$, **已知** $Y_1\sim\chi^2(n-1)$.

又 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$, **则由** χ^2 **分布定义知**:

$$Y_2=\frac{(\overline{X}-\mu)^2}{\sigma^2/n}\sim\chi^2(1).$$

再由 F **分布定义知**:

$$\frac{U_2}{(n-1)(\overline{X}-\mu)^2}=\frac{Y_1/(n-1)}{Y_2}\sim F(n-1,1).$$

例 6.8　总体 $X\sim N(0,9)$, $Y\sim N(0,9)$. $X_1, X_2,\cdots, X_9$ **与** $Y_1, Y_2,\cdots, Y_9$ **分别是来自** X、Y **的样本,两个样本相互独立,记**

$$T=\frac{X_1+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_9^2}},$$

则 T **的抽样分布是**(　　).

解　取　$\overline{X}=\dfrac{1}{9}(X_1+X_2+\cdots+X_9)$.

由 $X\sim N(0,3^2)$, **有**

$$\frac{\overline{X}-0}{3/\sqrt{9}}\sim N(0,1),$$

由 $Y\sim N(0,3^2)$, **有**

$$\frac{Y-0}{3}\sim N(0,1),$$

即　$$\frac{Y_i}{3}\sim N(0,1),\quad i=1,2,\cdots,9.$$

所以 $\sum\limits_{i=1}^{9}\dfrac{Y_i^2}{9}\sim\chi^2(9)$ **且与** $\overline{X}$ **相互独立,由** t **分布定义知**:

$$T = \frac{X_1 + \cdots + X_9}{Y_1^2 + Y_2^2 + \cdots + Y_9^2} = \frac{\overline{X}}{\frac{1}{9}\sum_{i=1}^{n}\frac{Y_1^2}{9}} \sim t(9) .$$

习 题 选 解

(6－1)　设样本的频数分布为:

X	0	1	2	3	4	5
频数	14	21	26	19	12	8

求样本均值 $\overline{x}$,样本方差 S^2,样本二阶中心矩 U_2 .

解　列表计算如下:

样本值 x_i	频数 f_i	分组和 $x_i f_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2 f_i$
0	14	0	-2.18	66.5336
1	21	21	-1.18	29.2404
2	26	52	-0.18	0.8424
3	19	57	0.82	12.7756
4	12	48	1.82	39.7488
5	8	40	2.82	63.6192
$\sum$	100	218	——	212.76

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = 2.18,$$

$$S^2 = \frac{\sum (x_i - \overline{x})^2 f_i}{99} = \frac{212.76}{99} = 2.149,$$

$$U_2 = \frac{212.76}{100} = 2.1276 .$$

(6-2) **设** $\overline{x}$, S_x^2 **为** $x_1, x_2, \cdots, x_n$ **的样本均值与样本方差.做数据变换**: $y_i = (x_i - a)/c\ (i = 1, 2, \cdots, n)$.**设** $\overline{y}$, S_y^2 **为** $y_1, y_2, \cdots, y_n$ **的样本均值与样本方差.证明**:

(1) $\overline{x} = a + c\overline{y}$;

(2) $S_x^2 = c^2 S_y^2$.

证 (1) **因为**

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - a)}{c}$$

$$= \frac{1}{nc}\sum_{i=1}^{n}(x_i - a) = \frac{1}{nc}\left(\sum_{i=1}^{n} x_i - na\right)$$

$$= \frac{1}{c}\overline{x} - \frac{a}{c},$$

所以 $\overline{x} = a + c\overline{y}$.

(2) **因为** $x_i = a + cy_i \quad (i = 1, 2, \cdots, n)$,

所以

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}[(a + cy_i) - (a + c\overline{y})]^2$$

$$= \frac{c^2}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = c^2 S_y^2.$$

(6-3) **设** $x_1, x_2, \cdots, x_n$ **是来自两点分布总体** X **的样本**, X **的分布为**

$$P\{X = 1\} = p,\ P\{X = 0\} = q \quad (q = 1 - p, 0 < p < 1).$$

求样本分布列.

解 **样本分布列为**

$$P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}$$

$$= P\{X_1 = x_1\}\cdot p\{X_2 = x_2\}\cdot\cdots\cdot P\{X_n = x_n\}$$

$$= (p^{x_1} q^{1-x_1})(p^{x_2} q^{1-x_2})\cdot\cdots\cdot(p^{x_n} q^{1-x_n})$$

$$= p^{\sum x_i} q^{\sum(1-x_i)} = p^{\sum\limits_{i=1}^{n} x_i} q^{n-\sum\limits_{i=1}^{n} x_i}.$$

(6-4) **设有** N **个产品,其中有** M **个次品**, $N-M$ **个正品.**

进行有放回抽样,定义 X_i 如下:

$$X_i=\begin{cases}1, & \text{第 } i \text{ 次取到次品;}\\ 0, & \text{第 } i \text{ 次取到正品.}\end{cases}$$

求样本 $X_1, X_2, \cdots, X_n$ 的分布.

解 因为是有放回抽样,则每次抽到次品的概率为 $p=\frac{M}{N}$;抽到正品的概率为 $q=\frac{N-M}{N}$.

所以
$$P(X=x_i)=p^{x_i}q^{1-x_i}\quad(i=1,2,\cdots,n),$$
$$P(X_1=x_1, X_2=x_2,\cdots,X_n=x_n)$$
$$=P(X_1=x_1)\cdot P(X_2=x_2)\cdot\cdots\cdot P(X_n=x_n)$$
$$=p^{\sum\limits_{i=1}^{n}x_i}q^{n-\sum\limits_{i=1}^{n}x_i}=\left(\frac{M}{N}\right)^{\sum\limits_{i=1}^{n}x_i}\left(\frac{N-M}{N}\right)^{n-\sum\limits_{i=1}^{n}x_i}.$$

(6-5) 设总体 X 服从泊松分布 $P(\lambda)$, $\overline{X}$ 是容量为 n 的样本的均值,求 $E(\overline{X})$, $D(\overline{X})$.

解 因为 $E(X)=\lambda$, $D(X)=\lambda$.而样本 $X_1, X_2,\cdots, X_n$ 相互独立且与总体 X 同分布,则

$$E(X_i)=\lambda,\ D(X_i)=\lambda,\quad i=1,2,\cdots,n.$$

所以
$$E(\overline{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\lambda.$$
$$D(\overline{X})=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{\lambda}{n}.$$

(6-6) 设总体 $X\sim N(40,5^2)$.

(1) 抽取容量为 36 的样本,求 $P\{38\le\overline{X}\le 43\}$;

(2) 抽取容量为 64 的样本,求 $P\{|\overline{X}-40|<1\}$;

(3) 样本容量 n 取多大时,才能使

$$P\{|\overline{X}-40|<1\}=0.95.$$

解 (1) 因为
$$\overline{X}\sim N\left(40,\frac{5^2}{36}\right),$$

所以
$$P\{38\le\overline{X}\le 43\}=\Phi\left(\frac{43-40}{5/6}\right)-\Phi\left(\frac{38-40}{5/6}\right)$$

$$= \Phi(3.6) - \Phi(-2.4) = \Phi(3.6) + \Phi(2.4) - 1$$
$$= 1 + 0.991\,8 - 1 = 0.991\,8.$$

(2) **因为** $\overline{X} \sim N\left(40, \dfrac{5^2}{64}\right)$,

所以 $P\{|\overline{X} - 40| < 1\} = P\{39 < \overline{X} < 41\}$

$$= \Phi\left(\frac{41-40}{5/8}\right) - \Phi\left(\frac{39-40}{5/8}\right) = \Phi(1.6) - \Phi(-1.6)$$
$$= 2\Phi(1.6) - 1 = 2\times 0.945\,2 - 1$$
$$= 0.890\,4.$$

(3) **因为** $\overline{x} \sim N\left(40, \dfrac{5^2}{n}\right)$, **则** $U = \dfrac{\overline{x} - 40}{5/\sqrt{n}} \sim N(0,1)$. **由**

$$0.95 = P\{|\overline{X} - 40| < 1\} = P\left\{|U| < \frac{1}{5/\sqrt{n}}\right\},$$

所以
$$P\left\{|U| > \frac{\sqrt{n}}{5}\right\} = 0.05.$$

由标准正态分布对称性可知:

$$P\left\{U > \frac{\sqrt{n}}{5}\right\} = 0.025.$$

则
$$P\left\{U < \frac{\sqrt{n}}{5}\right\} = 0.975.$$

由概率为 0.975 **查标准正态分布表得**

$$\frac{\sqrt{n}}{5} = 1.96, \quad \text{得 } n \approx 96.$$

(6-7) **设总体** $X \sim N(\mu_1, \sigma^2)$, **总体** $Y \sim N(\mu_2, \sigma^2)$. $X_1, X_2, \cdots, X_{n_1}$ **为来自总体** X **的样本**, $Y_1, Y_2, \cdots, Y_{n_2}$ **是来自总体** Y **的样本.设两个样本独立.** μ_1, μ_2 **已知.令**

$$\hat{\sigma}_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i - \mu_1)^2, \quad \hat{\sigma}_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i - \mu_2)^2.$$

求 $F = \dfrac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ **的抽样分布.**

解　令

$$\chi_1^2 = \frac{n_1 \sigma_1^2}{\sigma^2} = \frac{\sum_{i=1}^{n_1} (x_i - \mu_1)^2}{\sigma^2},$$

$$\chi_2^2 = \frac{\sum_{i=1}^{n_2} (y_i - \mu_2)^2}{\sigma^2} = \frac{n_2 \sigma_2^2}{\sigma^2},$$

则 $\chi_1^2 \sim \chi^2(n_1), \quad \chi_2^2 \sim \chi^2(n_2)$.

由 F 分布定义可知

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2} \sim F(n_1, n_2).$$

试 题 选 解

[6－1]　是非题

1. 设总体 X 的分布函数为 $F(x)$, $X_1, X_2, \cdots, X_n$ 为其样本, $F_n(x)$为 $X_1, X_2, \cdots, X_n$ 的经验分布函数, 则对任何固定的 x, $F_n(x)$不是统计量.　(　　)

答　错.

分析　因为 $F_n(x) = \dfrac{m(x)}{n}$, 其中 $m(x)$为样本值中不大于 x 的样本个数. 因此 $F_n(x)$为样本 $X_1, \cdots, X_n$ 的函数且不含未知参数, 应该是统计量.

2. 设样本 $X_1, X_2, \cdots, X_n$ 来自于总体 $N(\mu, \sigma^2)$, μ、σ^2 均未知, 则

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - (\overline{X} - \mu)^2$$

是统计量.　(　　)

答　正确.

分析 该样本函数表面上看含未知参数 μ.事实上

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-(\overline{X}-\mu)^2$$

$$=\frac{1}{n}\sum_{i=1}^{n}X_i^2-2\mu\overline{X}+\mu^2-(\overline{X}^2-2\mu\overline{X}+\mu^2)$$

$$=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2 .$$

并不含未知参数 μ.所以该样本函数是统计量 .

[6 - 2] **填空题**

1.设 $X_1,X_2,\cdots,X_n$ 是来自总体 $N(\mu,\sigma^2)$的样本,记

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad H_n^2=\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

则 $H_n^2/\sigma^2\sim$(　　),$H_n^2/[n(n-1)(\overline{X}-\mu)^2]\sim$(　　).

分析 (1) 由重要统计量的抽样分布知

$$H_n^2/\sigma^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1) .$$

(2) $H_n^2/[n(n-1)(\overline{X}-\mu)^2]$

$$=\frac{H_n^2}{\sigma^2}\Big/(n-1)\Big/\frac{(\overline{X}-\mu)^2}{\sigma^2/n}\sim F(n-1,1)$$

因为 $\dfrac{\overline{X}-\mu}{\sigma\sqrt{n}}\sim N(0,1)$,由 χ^2 分布定义知:

$$\left(\frac{\overline{X}-\mu}{\sigma\sqrt{n}}\right)^2=\frac{(\overline{X}-\mu)^2}{\sigma^2/n}\sim\chi^2(1) .$$

再由 F 分布定义知上述答案成立 .

2.设 $X_1,X_2,\cdots,X_n$ 为来自总体 $N(\mu,\sigma^2)$的样本,$\overline{X}$ 为样本均值,则

$$\frac{n(\overline{X}-\mu)^2}{S^2}\sim(\quad) .$$

分析 因为 $\dfrac{(\overline{X}-\mu)^2}{\sigma^2/n}\sim\chi^2(1)$,　$\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-$

1).

由 F 分布定义知:

$$\frac{n(\overline{X}-\mu)^2}{S^2}=\frac{(\overline{X}-\mu)^2/(\sigma^2/n)}{\frac{(n-1)S^2}{\sigma^2}\Big/(n-1)}\sim F(1,n-1).$$

3.设 $X_1,X_2,\cdots,X_n$ 是来自总体 $N(\mu,\sigma^2)$的样本,$\overline{X}$、S^2 分别为样本均值和样本方差,则$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim($ $)$.

答 $t(n-1)$.

直接由 t 分布的定义可得.

4.设样本 $X_1,X_2,\cdots,X_n$ 来自总体 $N(\mu,\sigma^2)$.$\overline{X}$、S^2 分别为样本均值和样本方差;又设 X_{n+1} 与 $X_1,X_2,\cdots,X_n$ 独立同分布,则 $\sqrt{\dfrac{n}{n+1}}\dfrac{X_{n+1}-\overline{X}}{S}\sim($ $)$.

分析 因为 $X_{n+1}\sim N(\mu,\sigma^2)$, $\overline{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$

所以 $X_{n+1}-\overline{X}\sim N\left(0,\sigma^2+\dfrac{\sigma^2}{n}\right)$.

令 $Y_1=\dfrac{X_{n+1}-\overline{X}}{\sigma\Big/\sqrt{\dfrac{n+1}{n}}}$,则 $Y_1\sim N(0,1)$.

又 $Y_2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

由 t 分布定义知:

$$\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\overline{X}}{S}=\frac{Y_1}{\sqrt{Y_2/(n-1)}}\sim t(n-1).$$

5.设样本的频数分布如下表

X	0	1	2	3	4
频数	1	3	2	1	2

则样本方差 $S^2=($ $)$.

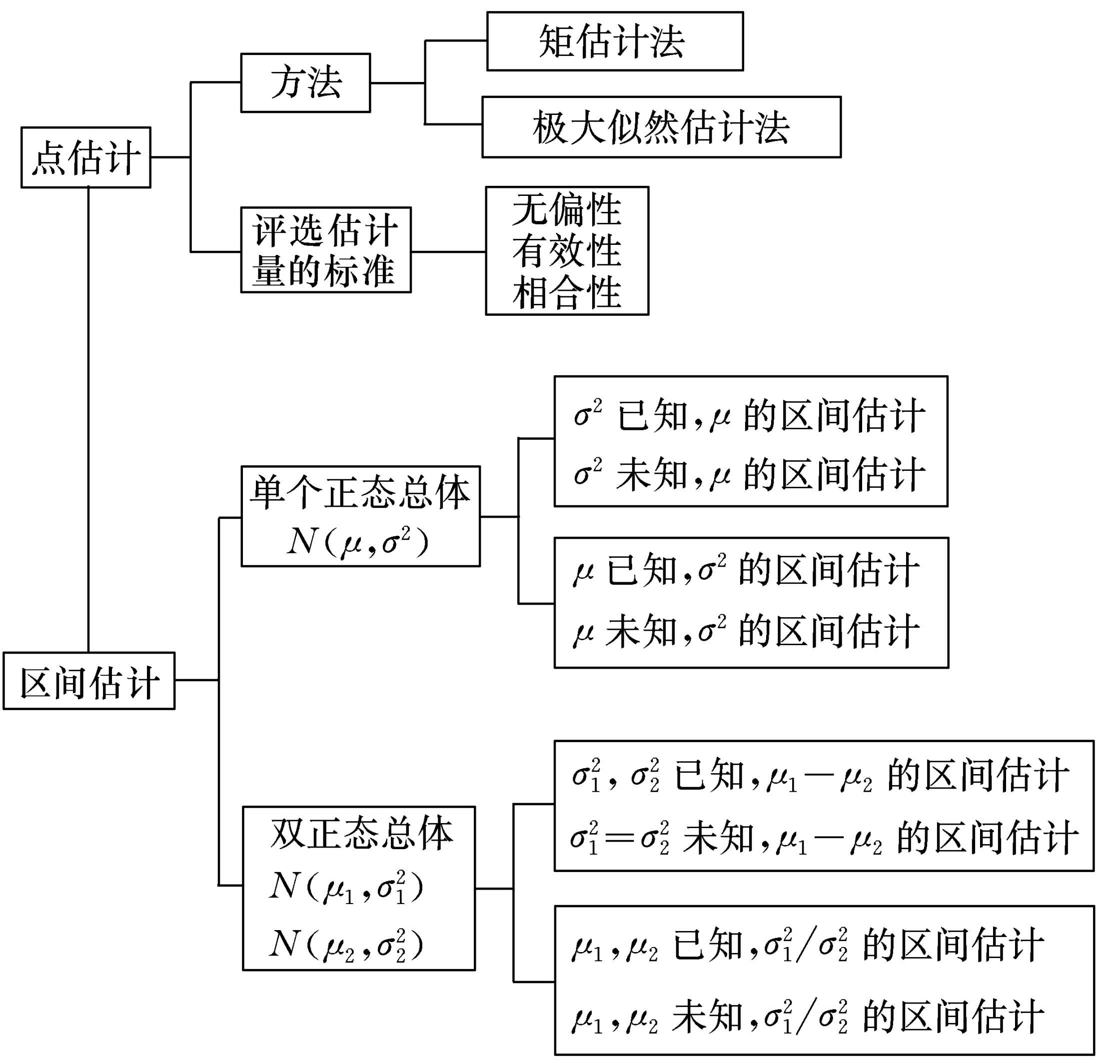
分析 $\overline{x}=\frac{1}{9}(0\times1+1\times3+2\times2+3\times1+4\times2)=2$,

$S^2=\frac{1}{8}[(-2)^2\times1+(-1)^2\times3+0^2\times2+1^2\times1+2^2\times2]=2$.

第七章　参 数 估 计

内 容 提 要

一、知识脉络



二、知识要点

(一) 点估计

1 . 定义

设 θ 为总体 X 的某个未知参数,用样本 $X_1, X_2, \cdots, X_n$ 的某个函数 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 去估计 θ,称 $\hat{\theta}$ 为 θ 的估计量;对于具体的样本值 $x_1, x_2, \cdots, x_n$,所得估计量 $\hat{\theta}$ 的值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 称为 θ 的点估计值 .

2 . 获得点估计量的两种常用方法

矩估计法　是用样本矩估计总体相应矩,从而得到未知参数估计量的方法 .常用的是一阶原点矩和二阶中心矩,即用

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} \quad 估 \quad E(X) = \mu,$$

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x}) \quad 估 \quad E[X - E(X)]^2 = \sigma^2 .$$

极大似然估计法　设 $x_1, x_2, \cdots, x_n$ 为来自总体 X 的样本 $X_1, X_2, \cdots, X_n$ 的一组观察值, X 的概率密度函数为 $f(x; \theta_1, \cdots, \theta_m)$, $\theta_1, \cdots, \theta_m$ 为待估参数,则称

$$L(x_1, x_2, \cdots, x_n, \theta_1, \cdots, \theta_m) = \prod_{i=1}^{n} f(x_i; \theta_1, \cdots, \theta_m)$$

为似然函数(X 为离散型,则 $L(x_1, x_2, \cdots, x_n; \theta_1, \cdots, \theta_m) = \prod_{i=1}^{n} P(x_i; \theta_1, \cdots, \theta_m)$) .若 $L(x_1, \cdots, x_n; \theta_1, \cdots, \theta_m)$ 在 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$ 达最大值,则称 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$ 为 $\theta_1, \theta_2, \cdots, \theta_m$ 的极大似然估计 .

(二) 估计量的评选标准

1 . 无偏性

设 $\hat{\theta}$ 是未知参数 θ 的一个估计量,若 $E(\hat{\theta}) = \theta$,则称 $\hat{\theta}$ 为 θ 的无偏估计量 .

2 . 有效性

设 $\hat{\theta}_1, \hat{\theta}_2$ 都是 θ 的无偏估计量,若 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$,则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效.如果对于给定的 n,在 θ 的所有无偏估计量中,$D(\hat{\theta})$ 的值最小,则称 $\hat{\theta}$ 是 θ 的有效估计量.

3.相合性

设 $\hat{\theta}$ 是 θ 的估计量,若对任意 $\varepsilon > 0$,有

$$\lim_{n \to \infty} P\{ |\hat{\theta}_n - \theta| < \varepsilon\} = 1,$$

则称 $\hat{\theta}$ 为 θ 的相合估计量.

(三) 区间估计

1.置信区间

设 θ 为总体 X 的一个未知参数,若对给定的 $\alpha(0 < \alpha < 1)$,由样本 $X_1, X_2, \cdots, X_n$ 确定两统计量 $\theta_1(X_1, \cdots, X_n), \theta_2(X_1, \cdots, X_n)$ 满足

$$P\{\theta_1 < \theta < \theta_2\} = 1 - \alpha,$$

则称区间 (θ_1, θ_2) 为未知参数 θ 的 $1 - \alpha$ 置信区间,θ_1 与 θ_2 分别为 θ 的置信下限和置信上限,$1 - \alpha$ 为置信度.

2.正态总体参数的区间估计表

正态总体参数的区间估计表见教材第 219～220 页。

要 点 范 例

一、未知参数的矩估计

解题思路　一般是先求总体的期望、方差等总体数字特征(矩),得到总体未知参数与总体数字特征的关系.而总体数字特征可由相应的样本数字特征(样本矩)来估计,由此得到未知参数的矩估计量.

例 7.1　总体 X 服从几何分布,分布列为

$$P\{X = k\} = (1 - p)^{k-1} p, \qquad k = 1, 2, \cdots.$$

其中 $0 < p < 1$.则 p 的矩估计是(　　).

答　$\dfrac{1}{\overline{X}}$.

分析　$\mu = E(X) = \sum\limits_{k=1}^{\infty} k(1-p)^{k-1}p$

$$= p\sum_{k=1}^{\infty} k\cdot q^{k-1} = p\cdot\frac{1}{(1-q)^2} = \frac{1}{p}.$$

而 $\hat{\mu} = \overline{X}$, 故 $\hat{p} = \dfrac{1}{\hat{\mu}} = \dfrac{1}{\overline{X}}$.(和函数利用幂级数逐项积分性质求得 .)

例 7 .2　总体 X 服从均匀分布 $U[a, b]$,则 a 的矩估计为(　　), b 的矩估计为(　　) .

答　$\overline{X} - \dfrac{3}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$,　$\overline{X} + \dfrac{3}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$.

分析　$\mu = E(X) = \dfrac{1}{2}(a+b)$,　得

$$a + b = 2\mu,$$

$$\sigma^2 = D(X) = \frac{1}{12}(b-a)^2, \quad 得$$

$$b - a = 2\ 3\sigma^2;$$

$$\hat{\mu} = \overline{X}, \quad \hat{\sigma}^2 = u_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

解方程组
$$\begin{cases} \hat{a} + \hat{b} = 2\overline{X}, \\ \hat{b} - \hat{a} = 2\ 3u_2, \end{cases}$$

得
$$\hat{a} = \overline{X} - 3u_2 = \overline{X} - \frac{3}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

$$\hat{b} = \overline{X} + 3u_2 = \overline{X} + \frac{3}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

例 7 .3　总体 X 的概率密度为

$$f(x_i;\ \phi) = \begin{cases} \dfrac{2\phi^2}{(\phi^2 - 1)x^3}, & 1 < x < \phi; \\ 0, & 其他 . \end{cases}$$

求 ϕ 的矩估计 .

解　$\mu = E(X) = \int_1^{\phi} x \frac{2\phi^2}{(\phi^2 - 1)x^3}\,dx = \frac{2\phi^2}{(\phi^2 - 1)}\int_1^{\phi} \frac{1}{x^2}\,dx$

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

得
$$\phi = \frac{\mu}{2 - \mu}.$$

而
$$\hat{\mu} = \overline{X}, \text{故 } \hat{\phi} = \frac{\hat{\mu}}{2 - \hat{\mu}} = \frac{\overline{X}}{2 - \overline{X}}.$$

二、未知参数的极大似然估计

解题思路　用极大似然估计法解题,可归纳成如下步骤.

(1) 用样本值构造似然函数

$$L(\phi) = f(x_1;\phi)\cdot f(x_2;\phi)\cdot\cdots\cdot f(x_n;\phi);$$

(2) 取 $Y = \ln[L(\phi)]$;

(3) 由 $\frac{dY}{d\phi} = 0$,得驻点 $\hat{\phi}$;

(4) 若 $\hat{\phi}$ 为 $L(\phi)$ 的最大值点,则 $\hat{\phi}$ 是 ϕ 的极大似然估计值. 将 $\hat{\phi}$ 中的样本值 $x_1, x_2, \cdots, x_n$ 换成样本 $X_1, X_2, \cdots, X_n$,则得极大似然估计量.

例 7.4　总体 X 服从几何分布:

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

其中 $0 < p < 1$, p 为未知参数,则 p 的极大似然估计是(　　).

答　$\frac{1}{\overline{X}}$.

分析　设 $x_1, x_2, \cdots, x_n$ 为样本值.

(1) 作似然函数

$$\begin{aligned}
L(p) &= P\{X = x_1, X = x_2, \cdots, X = x_n\} \\
&= P\{X = x_1\}\cdot p\{X = x_2\}\cdot\cdots\cdot p\{X = x_n\} \\
&= (1 - p)^{x_1 - 1} p\cdot(1 - p)^{x_2 - 1} p\cdot\cdots\cdot(1 - p)^{x_n - 1} p \\
&= (1 - p)^{\sum_{i=1}^{n} x_i - n} p^n.
\end{aligned}$$

(2) 取 $Y = \ln L(p) = \left(n - \sum_{i=1}^{n} x_i\right)\ln(1 - p) + n\ln p$.

(3) 令 $\dfrac{\mathrm{d}Y}{\mathrm{d}p} = -\dfrac{n - \sum\limits_{i=1}^{n} x_i}{1 - p} + \dfrac{n}{p} = 0$, 得惟一驻点 $\hat{p} = \dfrac{1}{\bar{x}}$.

(4) 因为 $\left.\dfrac{\mathrm{d}^2 Y}{\mathrm{d}p^2}\right|_{p=\hat{p}} < 0$, 所以 $\hat{p} = \dfrac{1}{\bar{x}}$ 为惟一极值点且为极大值点, 从而必为最大值点. 故 p 的极大似然估计值是 $\dfrac{1}{\bar{x}}$, p 的极大似然估计量是 $\hat{p} = \dfrac{1}{\bar{X}}$. 其中

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n).$$

例 7.5 设总体 X 的概率密度为

$$f(x;\phi,\lambda) = \begin{cases} \dfrac{1}{\lambda}\mathrm{e}^{-\frac{x-\phi}{\lambda}}, & x \geq \phi; \\ 0, & \text{其他}. \end{cases}$$

其中 $\lambda > 0$. 现抽得一批样本值 $x_1, x_2, \cdots, x_n$, 求未知参数 ϕ 和 λ 的极大似然估计值.

解 (1) 作似然函数

$$L(\phi,\lambda) = \prod_{i=1}^{n} f(x_i;\phi,\lambda)$$

$$= \begin{cases} \dfrac{1}{\lambda^n}\mathrm{e}^{-\frac{1}{\lambda}\sum\limits_{i=1}^{n}(x_i - \phi)}, & \min\{x_1, \cdots, x_n\} \geq \phi; \\ 0, & \text{其他}. \end{cases}$$

取 $\phi = \min\{x_1, \cdots, x_n\}$. 故

$$L(\phi,\lambda) = \begin{cases} \dfrac{1}{\lambda^n}\mathrm{e}^{-\frac{1}{\lambda}\sum\limits_{i=1}^{n}(x_i - \phi)}, & \forall x_i \geq \phi \quad (i = 1, \cdots, n); \\ 0, & \text{其他}. \end{cases}$$

(2) 当 $\forall x_i \geq \phi (i = 1, 2, \cdots, n)$ 时, 取

$$Y = \ln L(\phi,\lambda) = -\frac{1}{\lambda}\sum_{i=1}^{n}(x_i - \phi) - n\ln\lambda.$$

(3) 由
$$\frac{Y}{\lambda}=\frac{1}{\lambda^2}\sum_{i=1}^{n}(x_i-\varphi)-\frac{n}{\lambda}=0,$$
得 λ 的极大似然估计:
$$\lambda=\frac{1}{n}\sum_{i=1}^{n}(x_i-\varphi)=\overline{x}-\varphi,$$
其中 $\varphi=\min\{x_1,x_2,\cdots,x_n\}$.

三、估计量的无偏性、有效性、相合性

例 7.6 某车间生产的滚珠直径 $X\sim N(\mu,\sigma^2)$.随机抽取 6 个,测得直径(单位:mm)为 14.6,15.1,14.9,14.8,15.2,15.1.则

(1) 均值 μ 的无偏估计是(　　);

(2) 方差 σ^2 的无偏估计是(　　).

答　(1) 14.95;(2) 0.042 5.

分析　因为总体均值 μ 的无偏估计是样本均值 $\overline{x}$,方差 σ^2 的无偏估计是样本方差 S^2.

(1) $\overline{x}=15+\frac{1}{6}(-0.4+0.1-0.1-0.2+0.2+0.1)$
$=14.95$ (mm).

(2) $S^2=\frac{1}{6}(0.35^2+0.15^2+0.05^2+0.15^2+0.25^2+0.15^2)$
$=0.042\,5$.

例 7.7 总体 X 服从两点分布

X	1	0
P	p_1	q_1

总体 Y 服从两点分布

Y	1	0
P	p_2	q_2

其中 $0<p_1<1$, $p_1+q_1=1$; $0<p_2<1$, $p_2+q_2=1$. 样本 X_1, X_2, $\cdots$, X_{n_1} 来自总体 X, 样本 Y_1, Y_2, $\cdots$, Y_{n_2} 来自总体 Y, 两组样本独立. 则

(1) p_1-p_2 的无偏估计是(　　);

(2) 这一无偏估计的方差是(　　).

答　(1) $\overline{X}-\overline{Y}$; (2) $\frac{1}{n_1}p_1q_1+\frac{1}{n_2}p_2q_2$.

分析　(1) 因为

$$E(X)=p_1,\quad E(Y)=p_2.$$

$$\overline{X}=\frac{1}{n_1}\sum_{i=1}^{n_1}X_i,\quad \overline{Y}=\frac{1}{n_2}\sum_{i=1}^{n_2}Y_i,$$

$$E(\overline{X})=E(X)=p_1,\quad E(\overline{Y})=E(Y)=p_2.$$

所以

$$E(\overline{X}-\overline{Y})=p_1-p_2.$$

即 p_1-p_2 的无偏估计是 $\overline{X}-\overline{Y}$.

(2) 因为 $\overline{X}$ 与 $\overline{Y}$ 相互独立,

所以

$$\begin{aligned}D(\overline{X}-\overline{Y})&=D(\overline{X})+D(\overline{Y})\\&=\frac{1}{n_1^2}\sum_{i=1}^{n_1}D(X_i)+\frac{1}{n_2^2}\sum_{i=1}^{n_2}D(Y_i)\\&=\frac{1}{n_1}p_1q_1+\frac{1}{n_2}p_2q_2.\end{aligned}$$

例 7.8　总体 X 的均值 $E(X)=\mu$, 方差 $D(X)=\sigma^2(>0)$, 分别抽取容量为 10 和 15 的两组独立样本, $\overline{X}_1$, $\overline{X}_2$ 分别是这两组样本的均值. 令 $Y=a\overline{X}_1+b\overline{X}_2$, Y 是 μ 的无偏估计且使 $D(Y)$ 达到最小值. 则 $a=$(　　), $b=$(　　).

答　0.4, 0.6.

分析　因为 $E(\overline{X}_1)=E(\overline{X}_2)=E(X)=\mu$, 而

$$D(\overline{X}_1)=\frac{\sigma^2}{10},\quad D(\overline{X}_2)=\frac{\sigma^2}{15}.$$

又因为 Y 是 μ 的无偏估计.

$$E(Y)=E(a\overline{X}_1+b\overline{X}_2)=a\mu+b\mu=\mu\rightarrow a+b=1.$$

且 $a\overline{X}_1$ 与 $b\overline{X}_2$ 相互独立,有

$$D(Y)=D(a\overline{X}_1+b\overline{X}_2)=a^2D(\overline{X}_1)+b^2D(\overline{X}_2)$$

$$=\frac{a^2}{10}\sigma^2+\frac{b^2}{15}\sigma^2=\frac{3a^2+2(1-a)^2}{30}\sigma^2.$$

转化为求下述函数的最小值:

$$f(a)=3a^2+2(1-a)^2=5a^2-4a+2.$$

令 $f'(a)=10a-4=0$,得 $a=0.4$,进而得 $b=0.6$.

四、未知参数的区间估计

解题思路 在熟记正态总体各种抽样分布的基础上,注意归纳并熟练掌握区间估计的解题步骤.

(1) 根据问题类型,选取合适的样本函数 Y,然后根据 Y 所服从的抽样分布以及题中给定的置信度 $1-\alpha$,查相应的分布表,得到上下侧分位数(以下简称为临界值) a 和 b.

(2) 解不等式 $a<Y<b$,得参数 φ 的置信区间的置信上限 $Y_1=f_1(X_1,\cdots,X_n)$和置信下限 $Y_2=f_2(X_1,\cdots,X_n)$这两个统计量.

(3) 将样本值 $x_1,x_2,\cdots,x_n$ 代入上述两个统计量,得到参数 φ的置信区间.

例 7.9 某工厂生产滚珠的直径 $X\sim N(\mu,\sigma^2)$,从某日生产的产品中随机抽取 9 个,测得直径(单位:mm)为:14.6,14.7,15.1,14.9,14.8,15.0,15.1,15.2,14.8.在下面两种情形下,分别求直径均值 μ的置信度为 0.95 的置信区间:

(1) 若已知滚珠直径的标准差 $\sigma=0.15$ mm;

(2) 未知标准差 σ.

解 (1) 因已知方差,选取 $U=\dfrac{\overline{X}-\mu}{0.15/\sqrt{9}}\sim N(0,1)$.

由 $1-\alpha=0.95$ 得 $1-\dfrac{\alpha}{2}=0.975$,再查标准正态分布表得临

界值 $z_{\frac{\alpha}{2}} = 1.96$.

$$\bar{x} = 15 + \frac{1}{9}(-0.4 - 0.3 + 0.1 - 0.1$$
$$-0.2 + 0.1 + 0.2 - 0.2)$$
$$= 14.911.$$

置信上限:

$$\mu_2 = \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} = 14.911 + \frac{0.15}{\sqrt{9}} \times 1.96$$
$$= 14.911 + 0.098 = 15.009;$$

置信下限:

$$\mu_1 = \bar{x} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} = 14.911 - 0.098 = 14.813.$$

故滚珠直径均值 μ 的置信度为 0.95 的置信区间为

$$[14.813, 15.009].$$

(2) 因未知方差 σ, 选取 $T = \dfrac{\bar{X} - \mu}{s/\sqrt{n}}$, 则 $T \sim t(n-1)$.

由 $1 - \alpha = 0.95$ 得 $\dfrac{\alpha}{2} = 0.025$, 自由度 $n - 1 = 8$, 查 t 分布表得临界值 $t_{\frac{\alpha}{2}}(8) = 2.306$.

由样本值算得: $\bar{x} = 14.911$, $s = 0.203$.

置信上限:

$$\mu_2 = \bar{x} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) = 14.911 + \frac{0.203}{\sqrt{9}} \times 2.306$$
$$= 14.911 + 0.156 = 15.067;$$

置信下限:

$$\mu_2 = \bar{x} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) = 14.911 - 0.156 = 14.755.$$

故滚珠直径均值 μ 的置信度为 0.95 的置信区间为 $[14.755, 15.067]$.

例 7.10 设从两个正态总体 $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$ 中分别抽

取容量为 10 和 12 的样本,计算得到样本均值分别为 $\bar{x} = 20$, $\bar{y} = 24$,样本标准差分别为 $s_1 = 5$, $s_2 = 6$.求$(\mu_1 - \mu_2)$的置信度为 0 .95 的置信区间 .

分析　该题为两个正态总体在未知方差$(\sigma_1^2 = \sigma_2^2 = \sigma^2)$的条件下,求期望差 $\mu_1 - \mu_2$ 的区间估计 .选取

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w} \quad \frac{n_1 n_2}{n_1 + n_2}$$

其中
$$S_W = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} .$$

则
$$T \sim t(n_1 + n_2 - 2) .$$

解　由 $1 - \alpha = 0.95$ 得 $\frac{\alpha}{2} = 0.025$,自由度 $n_1 + n_2 - 2 = 20$,查 t 分布表得临界 $t_{\frac{\alpha}{2}}(20) = 2.086$.

计算得
$$S_W = \frac{9 \times 5^2 + 11 \times 6^2}{20} = 5.572,$$

$$\frac{1}{n_1} + \frac{1}{n_2} = \frac{1}{10} + \frac{1}{12} = 0.428 .$$

置信上限:

$$\begin{aligned} \phi_2 &= (\bar{x} - \bar{y}) + t_{\frac{\alpha}{2}}(20) S_W \quad \frac{1}{n_1} + \frac{1}{n_2} \\ &= -4 + 2.086 \times 5.572 \times 0.428 \\ &= -4 + 4.975 = 0.975; \end{aligned}$$

置信下限:

$$\begin{aligned} \phi_1 &= (\bar{x} - \bar{y}) - t_{\frac{\alpha}{2}}(20) S_W \quad \frac{1}{n_1} + \frac{1}{n_2} \\ &= -4 - 4.975 = -8.975 . \end{aligned}$$

故$(\mu_1 - \mu_2)$的置信度为 0 .95 的置信区间为$[-8.975, 0.975]$.

例 7 .11　设袋装糖果重 $X \sim N(\mu, \sigma^2)$,现随机抽取 16 袋,称得重量(单位: g)为: 506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496 .试求总体方差 σ^2 的置信度为

0.95的置信区间.

分析 该问题为单个正态总体的方差的区间估计.

选取 $Y=\dfrac{(n-1)S^2}{\sigma^2}$,则 $Y\sim\chi^2(n-1)$.

解 由 $1-\alpha=0.95$ 得 $\dfrac{\alpha}{2}=0.025$ 和 $1-\dfrac{\alpha}{2}=0.975$,自由度 $n-1=15$,查 χ^2 分布表得临界值:$\chi^2_{0.025}(15)=27.488$,$\chi^2_{0.975}(15)=6.262$.

计算得 $\bar{x}=503.75$,

$$(n-1)S^2=\sum_{i=1}^{16}(x_i-\bar{x})^2=577.$$

置信上限:

$$\sigma_2^2=\frac{(n-1)s^2}{\chi^2_{0.975}(15)}=\frac{577}{6.262}=92.143;$$

置信下限:

$$\sigma_1^2=\frac{(n-1)s^2}{\chi^2_{0.025}(15)}=\frac{577}{27.488}=20.991.$$

故总体方差 σ^2 的置信度为 0.95 的置信区间为 $[20.991, 92.143]$.

例 7.12 机器 A、B 生产的钢管内径 X、Y 分别服从 $N(\mu_1,\sigma_1^2)$和 $N(\mu_2,\sigma_2^2)$.从机器 A 生产的钢管中抽取 18 只,测得样本方差 $s_1^2=0.34$.抽取机器 B 生产的钢管 13 只,测得样本方差 $s_2^2=0.29$.两组样本相互独立.试求方差比 σ_1^2/σ_2^2 的置信度为 0.9 的置信区间.

分析 该问题为求两个正态总体方差比的区间估计.所以选取 $F=\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$.则 $F\sim F(n_1-1,n_2-1)$.

解 由 $1-\alpha=0.9$ 得 $\dfrac{\alpha}{2}=0.05$ 和 $1-\dfrac{\alpha}{2}=0.95$,再结合自由度$(n_1-1,n_2-1)=(17,12)$,查 F 分布表得临界值:$F_{0.05}(17,12)=2.59$;$F_{0.95}(17,12)=\dfrac{1}{F_{0.05}(12,17)}=\dfrac{1}{2.38}$.

置信上限

$$F_2 = \frac{S_1^2}{S_2^2} F_{0.95}^{-1}(17,12) = \frac{0.34}{0.29} \times 2.38 = 2.79 .$$

置信下限:

$$F_1 = \frac{S_1^2}{S_2^2} F_{0.05}^{-1}(17,12) = \frac{0.34}{0.29} \times \frac{1}{2.59} = 0.45 .$$

故 σ_1^2 / σ_2^2 的置信度为 0.9 的置信区间为 $[0.45, 2.79]$.

注　因为该区间包含 1,即置信下、上限分别在 1 左右.所以可认为两总体方差无明显差别.

习 题 选 解

(7-1)　使用测量仪器对同一值进行了 12 次独立测量,测量值为(单位: mm): 232.50, 232.48, 232.15, 232.53, 232.24, 232.30, 232.48, 232.05, 232.45, 232.60, 232.47, 232.30.用矩估计法估计测量值的真值与方差.

分析　正常情况下,测量值总是在真值附近波动.大数定律证明了测量值 X 的真值就是测量值的期望值 $E(X) = \mu$.而总体期望值的矩估计量就是样本平均值.测量值 X 的方差 $D(X) = \sigma^2$ 的矩估计量就是二阶中心矩.

解　
$$\begin{aligned}\hat{\mu} = \bar{x} &= 232.4 + \frac{1}{12}(0.1 + 0.08 - 0.25 \\ &\quad + 0.13 - 0.16 - 0.1 + 0.88 - 0.35 + 0.05 \\ &\quad + 0.2 + 0.07 - 0.1) \\ &= 232.379 .\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{12}(0.121^2 + 0.101^2 + 0.229^2 + 0.151^2 \\ &\quad + 0.139^2 + 0.079^2 + 0.101^2 + 0.329^2 + 0.071^2 \\ &\quad + 0.221^2 + 0.091^2 + 0.079^2)\end{aligned}$$

$= 0.026$.

(7－2) 设总体 X 具有概率密度

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

求 σ的极大似然估计.

解 设样本值为 $x_1, x_2, \cdots, x_n$.

作似然函数

$$L(\sigma) = \prod_{i=1}^{n} \frac{1}{2\sigma}e^{-\frac{|x_i|}{\sigma}} = (2\sigma)^{-n} \cdot e^{-\frac{1}{\sigma}\sum_{i=1}^{n}|x_i|}.$$

令 $$Y = \ln L(\sigma) = -n\ln(2\sigma) - \frac{1}{\sigma}\sum_{i=1}^{n}|x_i|.$$

令 $$\frac{dY}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}|x_i| = 0,$$

得 $$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}|x_i|.$$

(7－3) 设总体 X 服从均匀分布 $U(0, \phi)$,取得容量为 6 的样本值:1.3,0.6,1.7,2.2,0.3,1.1.求总体均值、总体方差的极大似然估计值.

解 因为 $X \sim U(0, \phi)$,则 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{\phi}, & \text{当 } 0 \leq x \leq \phi; \\ 0, & \text{其他}. \end{cases}$$

作似然函数

$$L(\theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

记 $$\hat{\theta} = \max\{x_1, x_2, \cdots, x_n\},$$

$$k = \min\{x_1, x_2, \cdots, x_n\}.$$

只有当 $0 \leq k \leq \hat{\theta} \leq \theta$ 时,

$$L(\theta) = \frac{1}{\theta^n}$$

否则 $L(\phi) = 0$.

$$L(\theta)=\frac{1}{\theta^n}\leq\frac{1}{\hat\theta^n}$$，只有当 $\theta=\hat\theta=\max\{x_1,x_2,\cdots,x_n\}$时，$L(\theta)$取到最大值 $L(\hat\theta)$.

因此 φ 的极大似然估计为 $\hat\theta=\max_i\{x_i\}$.

而 $\mu=E(X)=\frac{0+\theta}{2}=\frac{\theta}{2}$，$\sigma^2=D(X)=\frac{(\theta-0)^2}{12}=\frac{\theta^2}{12}$.

所以 $$\hat\mu=\frac{\hat\theta}{2}=\frac{1}{2}\max_i x_i=\frac{1}{2}\times 2.2=1.1,$$

$$\hat\sigma^2=\frac{\hat\theta^2}{12}=\frac{1}{12}\times 2.2^2=0.4033.$$

(7-4) 设总体 X 的概率密度为:

$$f(x)=\begin{cases}e^{-(x-\theta)}, & \text{当 } x\geq\phi;\\ 0, & \text{其他}.\end{cases}$$

用极大似然估计的定义证明 θ 的极大似然估计

$$\hat\theta=\min_i\{x_i\}.$$

证 作似然函数

$$L(\theta)=\prod_{i=1}^{n}f(x_i;\theta).$$

记 $$\hat\theta=\min\{x_1,x_2,\cdots,x_n\}=\min_i\{x_i\}.$$

只有当 $\hat\theta=\min_i\{x_i\}\geq\theta$时，

$$L(\theta)=e^{-(x_1-\theta)}e^{-(x_2-\theta)}\cdots e^{-(x_n-\theta)}=e^{-\sum_{i=1}^{n}x_i+n\theta}=e^{-n\bar{\theta}}\cdot e^{n\theta}$$

否则，$L(\theta)=0$.

显然 $L(\theta)$为 θ 的单调增加函数，而 $\theta\leq\hat\theta=\min_i\{x_i\}$.所以只有当 $\theta=\hat\theta=\min_i\{x_i\}$时，$L(\theta)$取到最大值 $L(\hat\theta)$.

根据定义，θ 的极大似然估计为

$$\hat\theta=\min_i\{x_i\}.$$

(7-5) 设总体 X 的均值 $E(X)=\mu$已知，方差 $D(X)=\sigma^2$未知.$X_1,X_2,\cdots,X_n$ 为一样本，证明:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

是 σ^2 的无偏估计 .

证　因为 x_i **相互独立且与总体** X **同分布 .故**

$$D(X_i) = E[(X_i - \mu)^2] = D(X) = \sigma^2,\ i = 1,2,\cdots,n .$$

所以
$$E(\hat{\sigma}^2) = E\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \ = \frac{1}{n}\sum_{i=1}^{n}E[(X_i - \mu)^2]$$
$$= \frac{1}{n}\sum_{i=1}^{n}D(X_i) = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}\cdot n\sigma^2$$
$$= \sigma^2 .$$

所以 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ **是方差 σ^2 的无偏估计 .**

(7 - 6)　设 X_1, X_2, X_3 **为总体** X **的样本 .证明:**

$$\hat{\mu}_1 = \frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{2}X_3,$$

$$\hat{\mu}_2 = \frac{2}{5}X_1 + \frac{1}{5}X_2 + \frac{2}{5}X_3$$

是总体均值 μ 的无偏估计,并判断其中哪一个估计较有效 .

证　设总体均值 $E(X) = \mu$, **方差** $D(X) = \sigma^2$.

因为 X_i **独立且与** X **同分布,则**

$$E(\hat{\mu}_1) = \frac{1}{6}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{2}E(X_3)$$
$$= \ \frac{1}{6} + \frac{1}{3} + \frac{1}{2}\ \ \mu = \mu .$$
$$E(\hat{\mu}_2) = \frac{2}{5}E(X_1) + \frac{1}{5}E(X_2) + \frac{2}{5}E(X_3)$$
$$= \ \frac{2}{5} + \frac{1}{5} + \frac{2}{5}\ \ \mu = \mu .$$

所以 $\hat{\mu}_1$、$\hat{\mu}_2$ **均为 μ 的无偏估计 .**

$$D(\hat{\mu}_1) = \frac{1}{36}D(X_1) + \frac{1}{9}D(X_2) + \frac{1}{4}D(X_3)$$

$$= \frac{1}{36} + \frac{1}{9} + \frac{1}{4} \quad \sigma^2 = \frac{7}{18}\sigma^2,$$

$$D(\hat{\mu}_2) = \frac{4}{25}D(X_1) + \frac{1}{25}D(X_2) + \frac{4}{25}D(X_3)$$

$$= \frac{4}{25} + \frac{1}{25} + \frac{4}{25} \quad \sigma^2 = \frac{9}{25}\sigma^2 .$$

因为 $D(\hat{\mu}_2) < D(\hat{\mu}_1)$,故 $\hat{\mu}_2$ 较 $\hat{\mu}_1$ 有效 .

(7 - 7) 设总体 $X \sim N(\mu_1, \sigma^2)$,总体 $Y \sim N(\mu_2, \sigma^2)$.样本 X_1, $X_2, \cdots, X_{n_1}$来自总体 X,样本 $Y_1, Y_2, \cdots, Y_{n_2}$来自总体 Y,两样本独立 .

(1) 求参数 $\mu_1 - \mu_2$ 的一个无偏估计 .

(2) 证明:

$$S_W^2 = \frac{1}{n_1 + n_2 - 2} \quad \sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$$

是 σ^2 的无偏估计,这里$\overline{X}$, $\overline{Y}$分别是两样本的均值 .

解 (1) 因为 $E(\overline{X}) = \frac{1}{n_1}\sum_{i=1}^{n_1} E(X_i) = \mu_1$,

$$E(\overline{Y}) = \frac{1}{n_2}\sum_{i=1}^{n_2} E(Y_i) = \mu_2,$$

故 $E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu_1 - \mu_2$.

因此 $\hat{\mu} = \overline{X} - \overline{Y}$ 是$(\mu_1 - \mu_2)$的无偏估计 .

(2) 因为 $\sum_{i=1}^{n_1}(X_i - \overline{X})^2 = (n_1 - 1)S_1^2$,

$$\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2 = (n_2 - 1)S_2^2 .$$

这里 S_1^2 为 $X_1, X_2, \cdots, X_{n_1}$ 的样本方差, S_2^2 为 $Y_1, Y_2, \cdots, Y_{n_2}$ 的样本方差 .则

$$E(S_1^2) = E(S_2^2) = \sigma^2 .$$

所以 $$E(S_W^2) = \frac{1}{n_1+n_2-2} E\left[\sum_{i=1}^{n_1}(X_i-\overline{X})^2 + \sum_{i=1}^{n_2}(Y_i-\overline{Y})^2\right]$$

$$= \frac{1}{n_1+n_2-2}\{E[(n_1-1)S_1^2] + E[(n_2-1)S_2^2]\}$$

$$= \frac{1}{n_1+n_2-2}[(n_1-1)\sigma^2 + (n_2-1)\sigma^2] = \sigma^2 .$$

故 S_W^2 是方差 σ^2 的无偏估计 .

(7-8) 设 $\hat{\theta}_1$ 及 $\hat{\theta}_2$ 是 θ 的两个独立的无偏估计量,且假定 $D(\hat{\theta}_1)=2D(\hat{\theta}_2)$,求常数 C_1 及 C_2,使 $\hat{\theta}=C_1\hat{\theta}_1+C_2\hat{\theta}_2$ 为 θ 的无偏估计,并使 $D(\hat{\theta})$ 达到最小 .

解 $$\theta = E(\hat{\theta}) = E[C_1\hat{\theta}_1 + C_2\hat{\theta}_2]$$

$$= C_1E(\hat{\theta}_1) + C_2E(\hat{\theta}_2) = (C_1+C_2)\theta .$$

得 $$C_1 + C_2 = 1 .$$

设 $D(\hat{\theta}_2)=\sigma^2$,则 $D(\hat{\theta}_1)=2\sigma^2$.又 $\hat{\theta}_1$,$\hat{\theta}_2$ 独立 .

所以 $$D(\hat{\theta}) = D[C_1\hat{\theta}_1 + C_2\hat{\theta}_2] = C_1^2D(\hat{\theta}_1) + C_2^2D(\hat{\theta}_2)$$

$$= 2C_1^2\sigma^2 + (1-C_1)^2\sigma^2 = (3C_1^2 - 2C_1 + 1)\sigma^2 .$$

令 $y=3C_1^2+2C_1+1$,显然 y 取最小值时 $D(\hat{\theta})$ 达到最小 .

令 $\frac{\mathrm{d}y}{\mathrm{d}C_1}=6C_1-2=0$ 得 $C_1=\frac{1}{3}$,而 $C_2=1-C_1=\frac{2}{3}$.

所以当 $C_1=\frac{1}{3}$, $C_2=\frac{2}{3}$ 时, $D(\hat{\theta})$ 达到最小 .

(7-9) 设灯泡厂生产的一大批灯泡的寿命 $X\sim N(\mu,\sigma^2)$,其中 μ、σ^2 未知 .今随机抽取 16 只灯泡进行寿命试验,测得寿命数据(单位: h)如下: 1502, 1480, 1485, 1511, 1514, 1527, 1603, 1480, 1532, 1508, 1490, 1470, 1520, 1505, 1485, 1540 .求这批灯泡平均寿命 μ 的置信度为 0.95 的置信区间 .

分析 该问题属于未知方差单个正态总体期望 μ 的区间估计,选取 $T=\frac{\overline{x}-\mu}{s/\sqrt{n}}$.

解 因方差未知，选取 $T=\dfrac{\overline{x}-\mu}{s/\sqrt{n}}$. 则 $T\sim t(n-1)$. 由 $1-\alpha=0.95$ 得 $\dfrac{\alpha}{2}=0.025$，自由度 $n-1=15$，查 t 分布表得临界值 $t_{\frac{\alpha}{2}}(15)=2.1315$.

计算得 $\overline{x}=1\,509.5$，$s=32.226$.

置信上限：

$$\hat{\phi}_2=\overline{x}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(15)=1\,509.5+\frac{32.226}{\sqrt{16}}\times 2.131\,5$$
$$=1\,509.5+17.172=1\,526.67;$$

置信下限：

$$\hat{\phi}_1=\overline{x}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(15)=1\,509.5-17.172=1\,492.33.$$

故期望 μ 的置信度为 0.95 的置信区间为 $[1\,492.33, 1\,526.67]$.

(7－10) 求上题灯泡寿命方差 σ^2 的置信度为 0.95 的置信区间.

解 选取 $\chi^2=\dfrac{(n-1)s^2}{\sigma^2}$，$\chi^2\sim\chi^2(n-1)$.

由 $1-\alpha=0.95$ 得 $\dfrac{\alpha}{2}=0.025$ 和 $1-\dfrac{\alpha}{2}=0.975$，自由度 $n-1=15$，查 χ^2 分布表得临界值：

$$\chi^2_{0.025}(15)=27.488,\quad \chi^2_{0.975}(15)=6.262.$$

计算得 $(n-1)s^2=15\times 32.226^2=15\,577.725$.

置信上限：

$$\hat{\sigma}_2^2=\frac{(n-1)s^2}{\chi^2_{0.975}(15)}=\frac{15\,577.725}{6.262}=2\,487.66;$$

置信下限：

$$\hat{\sigma}_1^2=\frac{(n-1)s^2}{\chi^2_{0.025}(15)}=\frac{15\,577.725}{27.488}=566.71.$$

故方差 σ^2 的置信度为 0.95 的置信区间为 $[566.71, 2\,487.66]$.

(7 - 11)　为了估计磷肥对农作物增产的作用,现选 20 块条件大致相同的土地.10 块不施磷肥,另外 10 块施磷肥,得亩产量(单位:kg)如下:

不施磷肥的:560,590,560,570,580,570,600,550,570,550.

施磷肥的:620,570,650,600,630,580,570,600,600,580.

设两类地亩产量均服从正态分布,且方差相同,试对施磷肥平均亩产量与不施磷肥平均亩产量之差做区间估计($\alpha = 0.05$).

解　选取 $T = \dfrac{(\overline{x} - \overline{y}) - (\mu_1 - \mu_2)}{S_W} \cdot \dfrac{n_1 n_2}{n_1 + n_2}$

其中
$$S_W = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

则 $T \sim t(n_1 + n_2 - 2)$.

由 $\alpha = 0.05$ 得 $\dfrac{\alpha}{2} = 0.025$,自由度 $n_1 + n_2 - 2 = 10 + 10 - 2 = 18$,查 t 分布表得临界值 $t_{\frac{\alpha}{2}}(18) = 2.1009$.

计算得 $\overline{x} = 600$, $\displaystyle\sum_{i=1}^{10}(x_i - \overline{x})^2 = 6\,400$.

$$\overline{y} = 570,\quad \sum_{i=1}^{10}(y_i - \overline{y})^2 = 2\,400.$$

$$S_W = \frac{6\,400 + 2400}{18} = 22.111.$$

$$\frac{1}{n_1} + \frac{1}{n_2} = \frac{1}{10} + \frac{1}{10} = 0.447.$$

置信上限:

$$\hat{\theta}_2 = (\overline{x} - \overline{y}) + S_W \frac{1}{n_1} + \frac{1}{n_2} t_{\frac{\alpha}{2}}(18)$$
$$= 30 + 22.111 \times 0.447 \times 2.1009$$
$$= 30 + 20.764 = 50.764.$$

置信下限:

$$\hat{\theta}_1 = (\overline{x} - \overline{y}) + S_W \frac{1}{n_1} + \frac{1}{n_2} t_{\frac{\alpha}{2}}(18)$$

$$= 30 - 20.764 = 9.236 .$$

故 $\mu_1 - \mu_2$ 的置信度为 0.95 的置信区间为 $[9.236, 50.764]$.

(7-12) 两位化验员 A、B 独立地对某种聚合物的含氮量用相同的方法分别做 10 次和 11 次测定.测定的方差分别为:

$$s_1^2 = 0.5419, \quad s_2^2 = 0.6065 .$$

设 A、B 两化验员测定值服从正态分布,其总体方差分别为 σ_1^2、σ_2^2. 求方差比 σ_1^2/σ_2^2 的置信度为 0.9 的置信区间.

解 选取 $F = \dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$,则 $F \sim F(n_1 - 1, n_2 - 1)$.由 $1 - \alpha = 0.9$ 得 $\dfrac{\alpha}{2} = 0.05$, $1 - \dfrac{\alpha}{2} = 0.95$.自由度为 $(n_1 - 1, n_2 - 1) = (9, 10)$,查 F 分布表得临界值:

$$F_{0.05}(9, 10) = 3.02,$$

$$F_{0.95}(9, 10) = \frac{1}{F_{0.05}(10, 9)} = \frac{1}{3.14} .$$

因为 $s_1^2 = 0.5419$, $s_2^2 = 0.6065$.

置信上限:

$$\hat{F}_2 = \frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{0.95}(9, 10)} = \frac{0.5419}{0.6065} \times 3.14 \approx 2.806;$$

置信下限:

$$\hat{F}_1 = \frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{0.05}(9, 10)} = \frac{0.5419}{0.6065} \times \frac{1}{3.02} \approx 0.296 .$$

故方差比 σ_1^2/σ_2^2 的置信度为 0.9 的置信区间为 $[0.296, 2.806]$.

试 题 选 解

[7-1] 是非题

1. 设 $\hat{\theta}$ 为 θ 的无偏估计,且对 θ 的任何无偏估计 θ',都有

$E(\hat{\theta}^2) \leq E(\hat{\theta}'^2)$.对一切 $\theta \in \Theta$,则 $\hat{\theta}$ 为 θ 的一致最小方差无偏估计.

答　正确.

分析　因为 $E(\hat{\theta}) = E(\hat{\theta}') = \theta$, $\quad E(\hat{\theta}^2) \leq E(\hat{\theta}'^2)$

而
$$D(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2 = E(\hat{\theta}^2) - \theta^2$$
$$D(\hat{\theta}') = E(\hat{\theta}'^2) - E(\hat{\theta}')^2 = E(\hat{\theta}'^2) - \theta^2 .$$
所以 $D(\hat{\theta}') \geq D(\hat{\theta})$.**其中** $\hat{\theta}'$**为** θ **的任一无偏估计.故** $\hat{\theta}$ **为** θ **的一致最小方差无偏估计.**

2.设 $X_1, X_2, \cdots, X_n$ **是来自正态总体** $N(\mu, \sigma^2)$**的样本,则** σ^2 **的极大似然估计必是** σ^2 **的无偏估计.**

答　不正确.

分析　σ^2 **的极大似然估计为** $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$,**它是二阶中心矩.**$E(\hat{\sigma}^2) = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$.**它不是** σ^2 **的无偏估计.**

3.设 $X_1, X_2, \cdots, X_n$ **是来自泊松分布** $P\{X = k\} = \dfrac{\varphi^k}{k!}\mathrm{e}^{-\varphi}(\varphi > 0, k = 0,1,2,\cdots)$ **的样本,则** φ **的矩估计是** φ **的一致估计量.**

答　正确.

分析　因为
$$\mu = E(X) = \sum_{k=0}^{+\infty} k \cdot \frac{\varphi^k}{k!}\mathrm{e}^{-\varphi} = \mathrm{e}^{-\varphi}\sum_{k=1}^{+\infty}\frac{\varphi^k}{(k-1)!} = \varphi .$$
所以 $\hat{\varphi} = \hat{\mu} = \overline{X}$ **是** φ **的矩估计量.**

$\overline{X}$ **是总体期望** μ **的一致估计量,即** $\overline{X}$ **是** φ **的一致估计量.**

4.矩估计一定是相合估计.(　　)

答　正确.

分析　由大数定律可证明样本 k **阶原点矩** $v_k = \dfrac{1}{n}\sum_{i=1}^{n} x_i^k$ **是总体** X **的** k **阶原点矩** $V_k = E(X^k)$ **的相合矩计,样本** k **阶中心矩**

$u_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$ 是总体 k 阶中心矩 $U_k = E[(X - E(X))^k]$ 的相合估计.所以矩估计一定是相合估计.

5.设 $\hat{\phi}$ 是 ϕ 的无偏估计,则 $g(\hat{\phi})$ 必是 $g(\phi)$ 的无偏估计.

答 不正确.

分析 这题缺 $g(x)$ 在 $x = \phi$ 处连续这个条件,所以不能肯定 $g(\hat{\phi})$ 必是 $g(\phi)$ 的无偏估计.

[7-2] 填空题

1.设总体 X 的密度函数为

$$f(x;\phi) = \begin{cases} \phi x^{\phi - 1}, & 0 < x < 1; \\ 0, & \text{其他}. \end{cases}$$

$X_1, X_2, \cdots, X_n$ 为其样本.则 ϕ 的矩估计 $\hat{\phi} = ($ $)$.

答 $\frac{\overline{X}}{1 - \overline{X}}$.

分析
$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x;\phi)\,\mathrm{d}x$$
$$= \int_0^1 x\phi \cdot x^{\phi - 1}\,\mathrm{d}x = \frac{\phi}{\phi + 1}.$$

而 $\hat{\mu} = \overline{X}$,故 $\hat{\mu} = \frac{\hat{\phi}}{\hat{\phi} + 1} = \overline{X}$.解得

$$\hat{\phi} = \frac{\overline{X}}{1 - \overline{X}}.$$

2.设总体 X 的密度函数为

$$f(x;\phi) = \begin{cases} (\phi + 1)x^{\phi}, & 0 < x < 1; \\ 0, & \text{其他}. \end{cases}$$

$X_1, X_2, \cdots, X_n$ 为其样本,则 ϕ 的矩估计 $\hat{\phi} = ($ $)$.

答 $\frac{2\overline{X} - 1}{1 - \overline{X}}$.

分析 $\mu = E(X) = \int_0^1 x(\phi + 1)x^{\phi}\,\mathrm{d}x = \frac{\phi + 1}{\phi + 2}$.

而 $\hat{\mu} = \overline{X}$,故 $\frac{\hat{\phi} + 1}{\hat{\phi} + 2} = \hat{\mu} = \overline{X}$.解得

$$\hat{\theta}=\frac{2\overline{X}-1}{1-\overline{X}}\text{ 或 }\frac{1-2\overline{X}}{\overline{X}-1}.$$

3 . 设总体 X 在 $[0,\theta]$ 上服从均匀分布, $X_1, X_2, \cdots, X_n$ 为其样本,则 θ 的矩估计 $\hat{\theta}=($　　$)$.

答　$2\overline{X}$.

分析　$\mu=E(X)=\frac{0+\theta}{2}=\frac{\theta}{2}$.

而 $\hat{\mu}=\overline{X}$, 故 $\hat{\theta}=2\hat{\mu}=2\overline{X}$.

4 . 设 X_1, X_2 是来自于 $N(0,\sigma^2)$ 的样本,则 σ^2 的置信水平为 $1-\alpha$ 的置信区间是(　) .

答　$\left(\frac{X_1^2+X_2^2}{\chi^2_{\frac{\alpha}{2}}(2)},\frac{X_1^2+X_2^2}{\chi^2_{1-\frac{\alpha}{2}}(2)}\right)$.

分析　因为期望 $\mu=0$ 已知, 选取 $\chi^2=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$.

由 $1-\alpha$ 得 $\frac{\alpha}{2}$ 和 $1-\frac{\alpha}{2}$ 的值, 自由度 $n=2$, 查 χ^2 分布表得临界值 $\chi^2_{\frac{\alpha}{2}}(2)$ 和 $\chi^2_{1-\frac{\alpha}{2}}(2)$.

$$\sum_{i=1}^{2}(X_i-\mu)^2=\sum_{i=1}^{2}X_i^2=X_1^2+X_2^2 .$$

置信上限　　$\hat{\sigma}_2^2=\frac{X_1^2+X_2^2}{\chi^2_{1-\frac{\alpha}{2}}(2)}$,

置信下限　　$\hat{\sigma}_1^2=\frac{X_1^2+X_2^2}{\chi^2_{\frac{\alpha}{2}}(2)}$,

所以置信区间为　　$\left(\frac{X_1^2+X_2^2}{\chi^2_{\frac{\alpha}{2}}(2)},\frac{X_1^2+X_2^2}{\chi^2_{1-\frac{\alpha}{2}}(2)}\right)$.

5 . 设 $X_1, X_2, \cdots, X_n$ 是来自 $[\theta,\theta+1]$ 上均匀分布的样本, θ 未知 .则 θ 的矩估计 $\hat{\theta}=($　　$)$.

答　$\overline{X}-\frac{1}{2}$.

分析　因为 $\mu = E(X) = \dfrac{\phi + (\phi + 1)}{2} = \phi + \dfrac{1}{2}$

所以 $\phi = \hat{\mu} - \dfrac{1}{2} = \overline{X} - \dfrac{1}{2}$.

[7－3]　**计算题**

1 . 设总体 X **服从** Γ **分布,其分布密度为**

$$f(x;\phi) = \begin{cases} \dfrac{1}{\phi^r \Gamma(r)} x^{r-1} e^{-\frac{x}{\phi}}, & x > 0; \\ 0, & x \leq 0 . \end{cases}$$

$r > 0$ **已知**, ϕ **未知**, $X_1, X_2, \cdots, X_n$ **为其样本** .

(1) **求** ϕ **的极大似然估计** ϕ;

(2) **证明** ϕ **为** ϕ **的无偏估计** .

(1) **解　作似然函数**:

$$L(\phi) = \prod_{i=1}^{n} f(x_i;\phi)$$

当 $\min\{x_1, x_2, \cdots, x_n\} > 0$ **时**,

$$L(\phi) = \prod_{i=1}^{n} \frac{1}{\phi^r \Gamma(r)} x_i^{r-1} e^{-\frac{x_i}{\phi}} = \Gamma(r)^{-n} \prod_{i=1}^{n} x_i^{\;r-1} \phi^{-nr} e^{-\frac{1}{\phi}\sum_{i=1}^{n} x_i} .$$

取 $Y = \ln L(\phi)$

$$= -n\ln\Gamma(r) + (r-1)\sum_{i=1}^{n}\ln x_i - nr\ln\phi - \frac{n}{\phi}\overline{x} .$$

令 $\dfrac{\mathrm{d}Y}{\mathrm{d}\phi} = -\dfrac{nr}{\phi} + \dfrac{n}{\phi^2}\overline{x} = 0$,**解得**

$$\phi = \frac{1}{r}\overline{X} \text{,为 } \phi \text{ 的极大似然估计 .}$$

(2) **证　因为** $E(\phi) = \dfrac{1}{r}E(\overline{X}) = \dfrac{1}{r}E(X)$

$$= \frac{1}{r}\int_0^{+\infty} x \cdot \frac{1}{\phi^r \Gamma(r)} x^{r-1} e^{-\frac{x}{\phi}} \mathrm{d}x$$

$$= \frac{\phi}{r} \cdot \frac{\Gamma(r+1)}{\Gamma(r)} = \phi .$$

所以 $\phi = \dfrac{1}{r}\overline{x}$ **为参数** ϕ **的无偏估计** .

2. **设** $X_1, X_2, \cdots, X_n$ **是来自均匀分布**

$$f(x;\phi)=\begin{cases}\dfrac{1}{\phi}, & 0\leq x\leq \phi;\\ 0, & \text{其他}\end{cases}$$

的样本,求 ϕ **的矩估计** ϕ **与极大似然估计** ϕ_L**,并问它们是否是** ϕ **的无偏估计**?

解 (1) **因为** $\mu = E(X)=\dfrac{0+\phi}{2}=\dfrac{\phi}{2}$,

所以 $\phi=2\mu=2\overline{X}$**为** ϕ **的矩估计量**.

而 $$E(\phi)=2E(\overline{X})=2E(X)=\phi.$$

所以 ϕ **为** ϕ **的无偏估计**.

(2) **作似然函数** $L(\phi)=\prod_{i=1}^{n} f(x_i;\phi)$

取 $$\phi_L=\max\{x_1, x_2, \cdots, x_n\}=\max_{1\leq i\leq n}\{x_i\},$$

$$k=\min\{x_1, x_2, \cdots, x_n\},$$

当 $0<k\leq \phi_L\leq \phi$ **时**, $L(\phi)=\dfrac{1}{\phi^n}$.

而 $L(\phi)=\dfrac{1}{\phi^n}$**为** ϕ **的单调下降函数,当** $\phi_0=\phi_L$ **时**, $L(\phi)$**取到最大值**.

所以 ϕ **的极大似然估计为** $\phi_L=\max\limits_{1\leq i\leq n}\{x_i\}$.

因为 ϕ_L **的分布函数为**

$$F(x)=\begin{cases}1, & x>\phi;\\ \left(\dfrac{x}{\phi}\right)^n, & 0<x\leq\phi;\\ 0, & x\leq 0.\end{cases}$$

则密度函数为

$$f(x)=\begin{cases}\dfrac{n}{\phi^n}x^{n-1}, & 0<x\leq\phi;\\ 0, & \text{其他}.\end{cases}$$

所以 $$E(\hat{\phi}_L)=\int_0^{\phi}x\cdot\frac{n}{\phi^n}x^{n-1}\mathrm{d}x=\frac{n}{n+1}\phi.$$

故 $\hat{\phi}_L$ 不是 ϕ 的无偏估计．

3．某烟厂生产了两种香烟，今独立、随机地抽取容量相同的烟叶标本，测其尼古丁含量（单位：mg），实验室分别做了 6 次测定，算得样本均值与方差分别为：

$$\bar{x}=25,\quad s_1^2=8;\quad \bar{y}=27,\quad s_2^2=11.$$

试给出两种烟叶尼古丁含量差的置信度为 95%的置信区间（这里假定尼古丁含量服从正态分布，且方差相等．）

解 选取 $T=\dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{S_W}\quad\dfrac{n_1n_2}{n_1+n_2}$，则 $T\sim t(n_1+n_2-2)$．由置信度 $1-\alpha=0.95$ 得 $\dfrac{\alpha}{2}=0.025$，自由度 $n_1+n_2-2=6+6-2=10$，查 t 分布表得临界值：

$$t_{\frac{\alpha}{2}}(10)=2.2281.$$

因为 $\bar{x}=25$，$\bar{y}=27$；$s_1^2=8$，$s_2^2=11$；$n_1=n_2=6$，

所以 $$S_W=\quad\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$=\quad\frac{5\times8+5\times11}{10}\approx3.082,$$

$$\frac{1}{n_1}+\frac{1}{n_2}=\quad\frac{1}{6}+\frac{1}{6}=0.577.$$

置信上限：

$$\hat{\phi}_2=(\bar{x}-\bar{y})+S_W\quad\frac{1}{n_1}+\frac{1}{n_2}\cdot t_{\frac{\alpha}{2}}(10)$$

$$=-2+3.082\times0.577\times2.2281$$

$$=-2+3.962=1.962;$$

置信下限：

$$\hat{\phi}_1=(\bar{x}-\bar{y})-S_W\quad\frac{1}{n_1}+\frac{1}{n_2}\cdot t_{\frac{\alpha}{2}}(10)$$

$$= -2-3.962 = -5.962.$$

所以两种烟叶尼古丁含量差的置信度为 95%的置信区间为 $[-5.962, 1.962]$.

4. (1) 设总体 X 的密度函数为

$$f(x,\phi) = \begin{cases} \phi x^{\phi-1}, & 0 < x < 1; \\ 0, & \text{其他}. \end{cases}$$

$\phi > 0$ 未知, $X_1, X_2, \cdots, X_n$ 为其样本.求 ϕ 的极大似然估计 ϕ.

(2) 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$的样本, 求 K 使得

$$\sigma^2 = \frac{1}{K}\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2 \text{ 为 } \sigma^2 \text{ 的无偏估计}.$$

解 (1) 作似然函数 $L(\phi) = \prod_{i=1}^{n} f(x_i; \phi)$.

当 " x_i 满足 $0 < x_i < 1$, $i = 1, 2, \cdots, n$ 时,

$$L(\phi) = \prod_{i=1}^{n} \phi x_i^{\phi-1} = \phi^n \left(\prod_{i=1}^{n} x_i\right)^{\phi-1}.$$

取 $Y = \ln L(\phi) = n\ln\phi + (\phi - 1)\sum_{i=1}^{n}\ln x_i$.

令 $\frac{\mathrm{d}Y}{\mathrm{d}\phi} = \frac{n}{\phi} + \sum_{i=1}^{n}\ln x_i = 0$, 得

$$\phi = -\frac{n}{\sum_{i=1}^{n}\ln x_i} \text{ 为 } \phi \text{ 的极大似然估计}.$$

(2) 因为 X_{i+1} 与 X_i 相互独立且 $X_{i+1} - X_i \sim N(0, 2\sigma^2)$而

$$E(X_{i+1} - X_i)^2 = D(X_{i+1} - X_i) = D(X_{i+1}) + D(X_i) = 2\sigma^2.$$

所以 $$E(\sigma^2) = \frac{1}{K}\sum_{i=1}^{n-1}E(X_{i+1} - X_i)^2 = \frac{2}{K}(n-1)\sigma^2 = \sigma^2,$$

必须 $$\frac{2}{K}(n-1) = 1, \text{得 } K = 2(n-1).$$

5. 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(0, \sigma^2)$的样本, σ^2 未知.

(1) 求 σ^2 的极大似然估计 σ^2.

(2) 证明 σ^2 为 σ^2 的无偏估计 .

解　(1)方法同前(略) .

答案: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

(2)

$$E(\sigma^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) = \frac{1}{n}\sum_{i=1}^{n} D(X_i) = \frac{1}{n}\cdot n\sigma^2 = \sigma^2$$

所以 σ^2 为 σ^2 的无偏估计 .

6 . 设总体 X 服从指数分布,密度函数为:

$$f(x;\varphi) = \begin{cases} \frac{1}{\varphi}e^{-\frac{x}{\varphi}}, & x > 0; \\ 0, & x \leq 0 . \end{cases}$$

其中 $\varphi > 0$ 未知, $x_1, x_2, \cdots, x_n$ 为其样本 .

(1) 求 φ 的极大似然估计 φ;

(2) 问 φ 是否为 φ 的无偏估计 ? 证明你的结论 .

解　(1) 作似然函数 $L(\varphi) = \prod_{i=1}^{n} f(x_i;\varphi)$.

当 " $x_i > 0, i = 1, 2, \cdots, n$ 时,

$$L(\varphi) = \prod_{i=1}^{n}\frac{1}{\varphi}e^{-\frac{x_i}{\varphi}} = \varphi^{-n}e^{-\frac{1}{\varphi}\sum_{i=1}^{n} x_i} .$$

取 $Y = \ln L(\varphi) = -n\ln(\varphi) - \frac{1}{\varphi}\sum_{i=1}^{n} x_i$.

令 $$\frac{dY}{d\varphi} = -\frac{n}{\varphi} + \frac{1}{\varphi^2}\sum_{i=1}^{n} x_i = 0,$$

解得　$\varphi = \overline{x}$ 为 φ 的极大似然估计 .

(2) 因为　$E(\varphi) = E\overline{X} = EX = \varphi$,

所以 $\varphi = \overline{X}$ 是 φ 的无偏估计 .

7 . 设总体 X 服从二项分布:

$$P\{X = k\} = C_N^k p^k q^{N-k}, \quad k = 0, 1, 2, \cdots, N .$$

其中 N 已知, $0 < p < 1$ 未知, $X_1, X_2, \cdots, X_n$ 为其样本 .

(1) 求 p 的极大似然估计 $\hat{p}$;

(2) $\hat{p}$ 是否为 p 的无偏估计？证明你的结论．

解　(1) 作似然函数

$$L(p)=\prod_{i=1}^{n}P(X=x_i)=\prod_{i=1}^{n}C_N^{x_i}p^{x_i}(1-p)^{N-x_i}$$

$$=\prod_{i=1}^{n}C_N^{x_i}\ p^{\sum_{i=1}^{n}x_i}(1-p)^{nN-\sum_{i=1}^{n}x_i}.$$

取 $Y=\ln L(p)$

$$=\ln\prod_{i=1}^{n}C_N^{x_i}+(\ln p)\sum_{i=1}^{n}x_i+\left(nN-\sum_{i=1}^{n}x_i\right)\ln(1-p).$$

令

$$\frac{\mathrm{d}Y}{\mathrm{d}p}=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{nN-\sum_{i=1}^{n}x_i}{1-p}=0,$$

解得　$\hat{p}=\dfrac{\overline{X}}{N}$为 p 的极大似然估计．

(2) 因为 $E(\hat{p})=\dfrac{1}{N}E(\overline{X})=\dfrac{1}{N}E(X)=\dfrac{1}{N}\cdot Np=p$．

所以 $\hat{p}=\dfrac{\overline{X}}{N}$为 p 的无偏估计．

8．某大学从来自 A、B 两市的新生中分别随机抽取 5 名与 6 名新生，测其身高(单位：cm)后算得 $\overline{x}=175.9$，$\overline{y}=172.0$；$s_1^2=11.3$，$s_2^2=9.1$．假设两市新生身高分别服从正态分布 $X\sim N(\mu_1,\sigma^2)$，$Y\sim N(\mu_2,\sigma^2)$，其中 σ^2 未知．试求 $\mu_1-\mu_2$ 的置信度为 0.95 的置信区间($t_{0.025}(9)=2.2622$，$t_{0.025}(11)=2.2010$)．

解　这是两正态总体均值差的区间估计，选取

$$T=\frac{(\overline{x}-\overline{y})-(\mu_1-\mu_2)}{S_W}\sqrt{\frac{n_1n_2}{n_1+n_2}},$$

则 $T\sim t(n_1+n_2-2)$，由 $1-\alpha=0.95$ 得$\dfrac{\alpha}{2}=0.025$，结合自由度 $n_1+n_2-2=5+6-2=9$，查 t 分布表得临界值

$$t_{\frac{\alpha}{2}}(9)=2.2622.$$

因为$\overline{x}=175.9$，$\overline{y}=172.0$；$s_1^2=11.3$，$s_2^2=9.1$；$n_1=5$，$n_2=6$．算得

$$S_W = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{4 \times 11.3 + 5 \times 9.1}{9} = 3.174\ 6 .$$

$$\frac{1}{n_1} + \frac{1}{n_2} = \frac{1}{5} + \frac{1}{6} = 0.605\ 6 .$$

置信上限:

$$\Phi_2 = (\overline{x} - \overline{y}) + S_W \frac{1}{n_1} + \frac{1}{n_2} t_{\frac{\alpha}{2}}(9)$$

$$= 3.9 + 3.174\ 6 \times 0.605\ 6 \times 2.262\ 2$$

$$= 3.9 + 4.349\ 2 \approx 8.249\ 2,$$

置信下限:

$$\Phi_1 = (\overline{x} - \overline{y}) - S_W \frac{1}{n_1} + \frac{1}{n_2} t_{\frac{\alpha}{2}}(9)$$

$$= 3.9 - 4.349\ 2 \approx -0.449\ 2 .$$

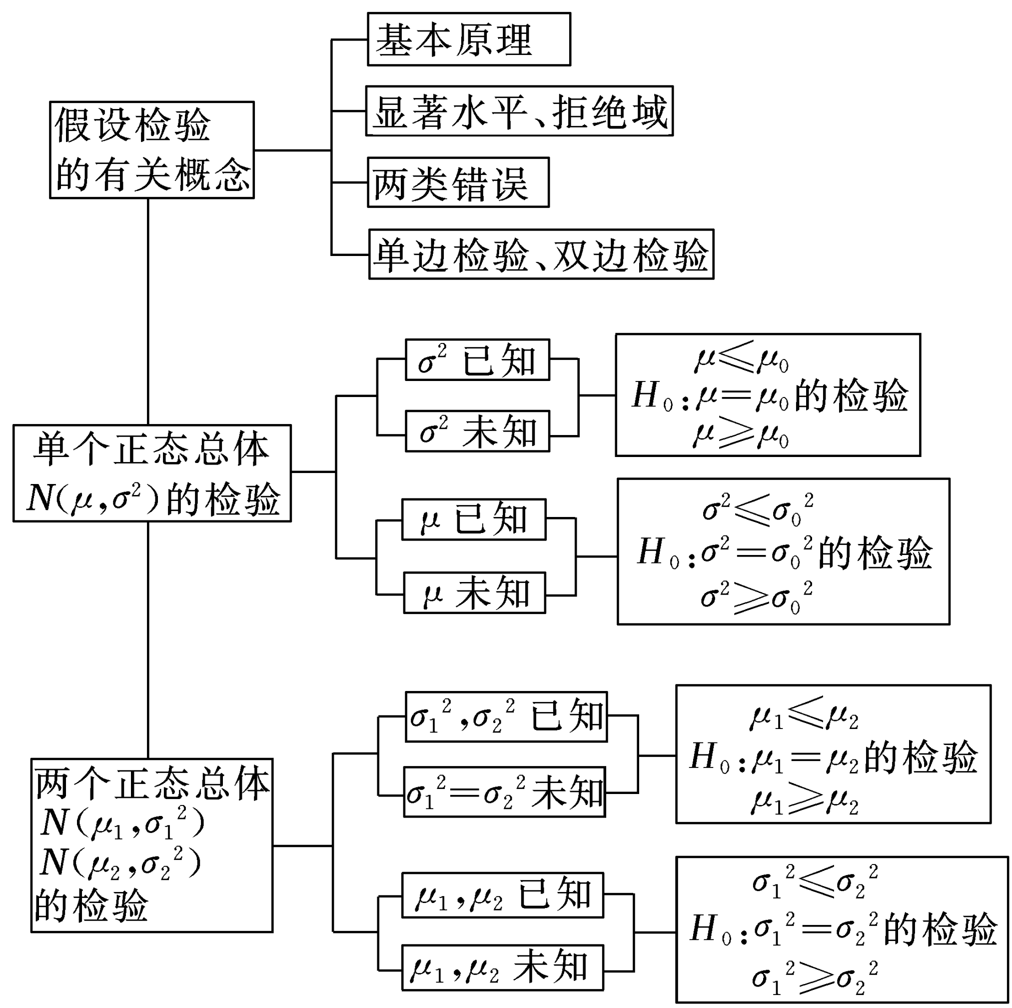
所以 $\mu_1 - \mu_2$ **的置信度为** 0.95 **的置信区间为** $[-0.449\ 2, 8.249\ 2]$.

注　几乎每届自考都有参数的区间估计题,但常与假设检验题并为一题.这类题型均放在第八章后面一并介绍.

第八章　假 设 检 验

内 容 提 要

一、知识脉络



二、知识要点

(一) 小概率原理

在一次试验中,发生概率很小(一般取概率 $\alpha=0.05$ 或 0.01)的事件几乎不可能发生.

(二) 基本思想

假设检验的基本步骤如下.

1. 提出假设

H_0——原假设;

H_1——备选假设(拒绝原假设后,准备选择的假设).

2. 推理分析

(1) 根据待检验问题的类型,选取相应的检验统计量,在原假设成立前提下,确定该检验统计量服从何种分布.

(2) 根据指定的小概率 α——显著水平(或检验水平)以及自由度查该分布的分布表得临界值.再由临界值可分出原假设 H_0 的拒绝域(小概率事件)和接受域.

3. 检验假设

根据一次试验后的样本值计算该检验统计量的值.若此值落在拒绝域,根据小概率原理,这是不大可能的,因此拒绝接受 H_0 而接受备选假设 H_1.若比值落在接受域,则无足够的理由拒绝 H_0,因此只能接受 H_0.

(三) 两类错误

1. 第一类错误(拒真错误)

拒真错误即以真为假的错误.犯这类错误的概率是检验统计量值落入拒绝域(或样本值落入由检验统计量定出的拒绝域) W 的概率,也就是显著水平 α.

$$\alpha=P\{(x_1,x_2,\cdots,x_n)\in W\mid H_0\ \text{成立}\}.$$

2. 第二类错误(取伪错误)

取伪错误即以假当真的错误.犯这类错误的概率是 β(计算较复杂,注意不能错误地认为是 $1-\alpha$).

$$\beta = p\{(x_1, x_2, \cdots, x_n) \notin W \mid H_1 \text{ 成立}\}.$$

3. 两类错误的关系

(1) 样本容量 n 固定时,α 减小,则 β 增加;β 减小时,则 α 增加.

(2) 要同时降低 α、β,只有通过增加样本容量 n 来解决.

(四) 正态总体均值的假设检验表

H_0	H_1	条件	检验统计量	拒绝域
$\mu = \mu_0$	$\mu \neq \mu_0$	σ^2 已知	$u = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$\lvert u \rvert > z_{\frac{\alpha}{2}}$
$\mu \leq \mu_0$	$\mu > \mu_0$	σ^2 已知		$u > z_\alpha$
$\mu \geq \mu_0$	$\mu < \mu_0$	σ^2 已知		$u < z_\alpha$
$\mu = \mu_0$	$\mu \neq \mu_0$	σ^2 未知	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$\lvert t \rvert > t_{\frac{\alpha}{2}}(n-1)$
$\mu \leq \mu_0$	$\mu > \mu_0$	σ^2 未知		$t > t_\alpha(n-1)$
$\mu \geq \mu_0$	$\mu < \mu_0$	σ^2 未知		$t < -t_\alpha(n-1)$

$\mu_1=\mu_2$	$\mu_1\neq\mu_2$	$\sigma_1^2=\sigma_2^2$ 未知	$t=\frac{(\bar{x}-\bar{y})}{S_W}\ \frac{n_1 n_2}{n_1+n_2}$ 其中 $S_W^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$	$\lvert t\rvert > t_{\frac{\alpha}{2}}(n_1+n_2-2)$
$\mu_1\leq\mu_2$	$\mu_1>\mu_2$	$\sigma_1^2=\sigma_2^2$ 未知		$t>t_\alpha(n_1+n_2-2)$
$\mu_1\geq\mu_2$	$\mu_1<\mu_2$	$\sigma_1^2=\sigma_2^2$ 未知		$t<t_\alpha(n_1+n_2-2)$

(五) 正态总体方差的假设检验表

H_0	H_1	条件	检验统计量	拒绝域
$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	μ未知	$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2<\chi^2_{1-\frac{\alpha}{2}}(n-1)$ $\chi^2>\chi^2_{\frac{\alpha}{2}}(n-1)$
$\sigma^2\leq\sigma_0^2$	$\sigma^2>\sigma_0^2$	μ未知		$\chi^2>\chi^2_\alpha(n-1)$
$\sigma^2\geq\sigma_0^2$	$\sigma^2<\sigma_0^2$	μ未知		$\chi^2<\chi^2_{1-\alpha}(n-1)$
$\sigma_1^2=\sigma_2^2$	$\sigma_1^2\neq\sigma_2^2$	μ_1,μ_2 未知	$F=S^2_{大}/S^2_{小}$	$F>F_{\frac{\alpha}{2}}(n_{大}-1,n_{小}-1)$
$\sigma_1^2\leq\sigma_2^2$	$\sigma_1^2>\sigma_2^2$	μ_1,μ_2 未知	$F=S_1^2/S_2^2$	$F>F_\alpha(n_1-1,n_2-1)$
$\sigma_1^2\geq\sigma_2^2$	$\sigma_1^2<\sigma_2^2$	μ_1,μ_2 未知	$F=S_2^2/S_1^2$	$F>F_\alpha(n_2-1,n_1-1)$

要 点 范 例

一、假设检验的基本概念

例 8.1 设总体 $X\sim N(\mu,\sigma^2)$, μ,σ^2 未知, $X_1,X_2,\cdots,X_n$ 是来自该总体的样本.记

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad Q=\sum_{i=1}^{n}(X_i-\overline{X})^2 .$$

则对于假设检验(显著水平 α)

$$H_0: \mu = \mu_0 \backslash \quad H_1: \mu \neq \mu_0 .$$

使用的检验统计量为(　　)(用 $\overline{X}$、Q 表示);其拒绝域为(　　).

解　这是单个正态总体均值 μ 的双侧假设检验(未知方差),应选取统计量 $T = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$,其中 $s^2 = \dfrac{Q}{n-1}$,则 $s = \sqrt{\dfrac{Q}{n-1}}$.

因为 $T \sim t(n-1)$,且是双侧检验.所以由显著水平 α 及自由度 $n-1$ 查 t 分布表可得临界值 $t_{\frac{\alpha}{2}}(n-1)$,使

$$P\{|T| > t_{\frac{\alpha}{2}}(n-1)\} = \alpha .$$

所以检验统计量为

$$T = \frac{\overline{X} - \mu_0}{\sqrt{\dfrac{Q}{n(n-1)}}} \text{ 或 } \frac{(\overline{X} - \mu_0)\sqrt{n(n-1)}}{\sqrt{Q}},$$

拒绝域为 $W = \{|T| > t_{\frac{\alpha}{2}}(n-1)\}$.

例 8.2　设总体 $X \sim N(\mu, \sigma^2)$,μ、σ^2 未知.$X_1, X_2, \cdots, X_n$ 是来自该总体的样本,样本均值 $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$,样本方差

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{ (显著水平 } \alpha) .$$

(1) 对于假设检验 $H_0: \mu \leq 1 \backslash \quad H_1: \mu > 1$ 的检验统计量为(　　),拒绝域 $W_1 =$(　　).

(2) 对于 $H_0: \mu \geq 1 \backslash \quad H_1: \mu < 1$,拒绝域 $W_2 =$(　　).

分析　这是单个正态总体均值 μ 的单侧检验(未知方差)问题.应选取检验统计量

$$T = \frac{\overline{X} - 1}{s/\sqrt{n}} \text{ 或 } T = \frac{(\overline{X} - 1)\sqrt{n}}{s}, \quad T \sim t(n-1) .$$

解　(1) 由显著水平 α 和自由度 $n-1$ 查 t 分布表得临界值 $t_\alpha(n-1)$,使

$$P\{T > t_\alpha(n-1)\} = \alpha.$$

所以拒绝域 $W_1 = \{T > t_\alpha(n-1)\}$.

(2) 由显著水平 α 和自由度 $n-1$ 查 t 分布表得临界值 $t_\alpha(n-1)$,使

$$P\{T < -t_\alpha(n-1)\} = P\{T > t_\alpha(n-1)\} = \alpha.$$

所以拒绝域 $W_2 = \{T < -t_\alpha(n-1)\}$.

例 8.3 总体 $X \sim N(10, \sigma^2)$, $X_1, X_2, \cdots, X_{16}$ 是来自 X 的样本,对于检验

$$H_0: \sigma^2 = 0.25 \quad H_1: \sigma^2 \neq 0.25$$

其检验统计量为(　　),拒绝域 W 为(　　).

解 这是单正态总体方差(已知均值 $\mu = 10$)的双侧检验问题.应选取检验统计量 $\chi^2 = \dfrac{\sum_{i=1}^{16}(X_i - 10)^2}{0.25}$,则 $\chi^2 \sim \chi^2(n)$ 即 $\chi^2(16)$.

由显著水平 α 得 $\dfrac{\alpha}{2}$ 和 $1-\dfrac{\alpha}{2}$,再结合自由度 $n = 16$ 查 χ^2 分布表得上临界值 $\chi^2_{\frac{\alpha}{2}}(16)$ 和下临界值 $\chi^2_{1-\frac{\alpha}{2}}(16)$.

$$P\{\chi^2_{1-\frac{\alpha}{2}}(16) < \chi^2 < \chi^2_{\frac{\alpha}{2}}(16)\} = 1 - \alpha.$$

所以
$$P\{\chi^2 > \chi^2_{\frac{\alpha}{2}}(16)\} = P\{\chi^2 < \chi^2_{1-\frac{\alpha}{2}}(16)\} = \frac{\alpha}{2}.$$

拒绝域 $W = \{\chi^2 < \chi^2_{1-\frac{\alpha}{2}}(16)\} \cup \{\chi^2 > \chi^2_{\frac{\alpha}{2}}(16)\}$.

例 8.4 总体 $X \sim N(\mu, \sigma^2)$,其中 μ 未知样本,$X_1, X_2, \cdots, X_n$ 来自于总体 X.样本均值

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

样本方差

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

对于检验问题(显著水平 α),

$$H_0: \sigma^2 \geq 9 \backslash \quad H_1: \sigma^2 < 9$$

选取检验统计量为(　　),其拒绝域 W 为(　　).

解　这是单正态总体(未知均值 μ)方差的单侧检验问题,应选取检验统计量

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(n-1)s^2}{9},$$

它服从自由度为 $n-1$ 的 χ^2 分布.

因为是单侧(上侧)检验,由显著水平 α得 $1-\alpha$.再结合自由度 $n-1$查 χ^2 分布表得临界值 $\chi^2_{1-\alpha}$,使

$$P\{\chi^2 > \chi^2_{1-\alpha}(n-1)\} = 1-\alpha$$

从而

$$P\{\chi^2 < \chi^2_{1-\alpha}(n-1)\} = \alpha$$

所以拒绝域 $W = \{\chi^2 < \chi^2_{1-\alpha}(n-1)\}$.

例 8.5　总体 $X \sim N(\mu_1, 0.25)$,总体 $Y \sim N(\mu_2, 0.16)$.样本 $X_1, X_2, \cdots, X_{16}$ 来自总体 X,样本 $Y_1, Y_2, \cdots, Y_9$ 来自总体 Y,两样本独立,对于检验

$$H_0: \mu_1 = \mu_2 \backslash \quad H_1: \mu_1 \neq \mu_2$$

应选取检验统计量为(　　),其拒绝域 W 为(　　).

解　这是两正态总体已知方差对均值差的双侧检验问题,应选取检验统计量

$$U = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{0.25}{16}+\frac{0.16}{9}}}$$

$$= \frac{24(\overline{X}-\overline{Y})}{19.24}.$$

因为$\overline{X}$、$\overline{Y}$独立,所以在 H_0 下: $\mu_1 - \mu_2 = 0$, $E(\overline{X}-\overline{Y}) = 0$,

$$D = (\overline{X}-\overline{Y}) = D(\overline{X}) + D(\overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2},$$

所以

$$U \sim N(0,1).$$

由显著水平 α 得 $1-\frac{\alpha}{2}$,查标准正态分布表得临界值 $z_{\frac{\alpha}{2}}$,使

$$P\{|U|>z_{\frac{\alpha}{2}}\}=\alpha .$$

所以拒绝域 W 为 $\left|\frac{24(\overline{X}-\overline{Y})}{19.24}\right|>z_{\frac{\alpha}{2}}$.

例 8.6 总体 $X\sim N(\mu_1,\sigma_1^2)$,总体 $Y\sim N(\mu_2,\sigma_2^2)$,其中 μ_1, μ_2 未知.样本 $X_1, X_2,\cdots, X_{n_1}$ 来自于总体 X,样本 $Y_1, Y_2,\cdots, Y_{n_2}$ 来自于总体 Y.两样本独立.设两样本的方差分别为 s_1^2, s_2^2.对于检验(显著水平为 α)

$$H_0:\sigma_1^2=\sigma_2^2\setminus\ H_1:\sigma_1^2\neq\sigma_2^2$$

其检验统计量 $F=($ $)$,拒绝域 W 为( ).

解 这是两正态总体(未知期望)的方差齐性的检验问题,应选取检验统计量

$$F=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}=\frac{s_1^2}{s_2^2} .$$

则 $F\sim F(n_1-1, n_2-1)$.

因为是双侧检验,由显著水平 α 得 $\frac{\alpha}{2}$ 和 $1-\frac{\alpha}{2}$,再结合两自由度 (n_1-1, n_2-1) 查 F 分布表得临界值

$$F_{\frac{\alpha}{2}}(n_1-1, n_2-1),\quad F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1),$$

使 $P\{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)<F<F_{\frac{\alpha}{2}}(n_1-1, n_2-1)\}=1-\alpha .$

所以拒绝域

$$W=\{F<F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)\}\cup\{F>F_{\frac{\alpha}{2}}(n_1-1, n_2-1)\} .$$

例 8.7 设原假设为 H_0,备选假设为 H_1,拒绝域为 W,取得的样本为 $X_1, X_2,\cdots, X_n$,则假设检验犯第一类错误的概率 $\alpha=$ ( ),犯第二类错误的概率为( ).

解 只需理解并记忆如下结论:

$$\alpha=P\{(x_1, x_2,\cdots, x_n)\in W\mid H_0\text{ 成立}\},$$

$$\beta = P\{(x_1, x_2, \cdots, x_n) \notin W \mid H_1 \text{ 成立}\}.$$

例 8.8 某厂生产的一种螺钉,标准要求长度是 68 mm,实际生产的产品,其长度服从 $N(\mu, 3.6^2)$.考虑检验问题

$$H_0: \mu = 68 \leftrightarrow H_1: \mu \neq 68.$$

设 $\overline{X}$ 为样本均值.按下列方式进行假设检验:

当 $|\overline{X} - 68| > 1$ 时,拒绝假设 H_0;

当 $|\overline{X} - 68| \leq 1$ 时,接受假设 H_0.

(1) 当样本容量 $n = 36$ 时,求犯第一类错误的概率 α;

(2) 当 $n = 64$ 时,求犯第一类错误的概率 α;

(3) 当 H_0 不成立(设 $\mu = 70$), $n = 64$ 时,按上述检验法,求犯第二类错误的概率 β.

分析 该问题是已知方差的单正态总体均值的双侧检验问题.应取统计量 $U = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$.

解 (1) 在假设 H_0 成立的前提下, $n = 36$,所以取统计量

$$U = \frac{\overline{X} - 68}{3.6/\sqrt{36}} = \frac{\overline{X} - 68}{0.6},$$

则 $U \sim N(0,1)$.拒绝域

$$W = \{(X_1, X_2, \cdots, X_n) \mid |\overline{X} - 68| > 1\}$$

$$= \left\{\frac{|\overline{X} - 68|}{0.6} > \frac{1}{0.6}\right\} = \{|U| > 1.67\}$$

所以

$$\begin{aligned}\alpha &= p\{|U| > 1.67 \mid H_0 \text{ 成立}\} = 2P\{U > 1.67 \mid H_0 \text{ 成立}\} \\ &= 2[1 - \Phi(1.67)] = 2(1 - 0.9525) \\ &= 0.0950.\end{aligned}$$

(2) 因为 $n = 64$,则 $U = \dfrac{\overline{X} - 68}{3.6/\sqrt{64}} = \dfrac{\overline{X} - 68}{0.45}$.拒绝域

$$W = \left\{\frac{|\overline{X} - 68|}{0.45} > \frac{1}{0.45}\right\} = \{|U| > 2.22\}$$

所以 $\alpha = P\{|U| > 2.22 | H_0 \text{成立}\}$

$= 2P\{U > 2.22 | H_0 \text{成立}\} = 2[1 - \Phi(2.22)]$

$= 2(1 - 0.9868) = 0.0264$.

(3) 因为 $n = 64$, $\mu = 70$.则 $\overline{X} \sim N\left(70, \frac{3.6^2}{64}\right)$ 即 $N(70, 0.45^2)$.

$$\beta = p\{(X_1, X_2, \cdots, X_n) \notin W | \mu = 70\}$$

$$= p\{|\overline{X} - 68| \leq 1 | \mu = 70\} = p\{67 \leq \overline{X} \leq 69 | \mu = 70\}$$

$$= \Phi\left(\frac{69 - 70}{0.45}\right) - \Phi\left(\frac{67 - 70}{0.45}\right) = \Phi(-2.22) - \Phi(-6.67)$$

$$= \Phi(6.67) - \Phi(2.22) = 1 - 0.9868 = 0.0132.$$

二、正态总体均值的假设检验

例 8.9 某种产品重量(单位:g) $X \sim N(12, 1)$,更新设备后,从新生产的产品中,随机抽取 100 个,测得样本均值 $\overline{x} = 12.5$(g).若方差没有变化,问设备更新后,产品的平均重量是否有显著变化($\alpha = 0.1$)?

分析 这是单正态总体(已知方差)均值的双侧检验问题,故选取检验统计量

$$U = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}},$$

解 提出假设

$$H_0: \mu = 12; \quad H_1: \mu \neq 12.$$

推理分析

选取 $U = \dfrac{\overline{x} - 12}{1/\sqrt{100}}$, $U \sim N(0, 1)$,由 $\alpha = 0.1$ 得 $1 - \dfrac{\alpha}{2} = 0.95$,查标准正态分布表得临界值

$$z_{\frac{\alpha}{2}} = 1.645.$$

检验判断

因为 $\bar{x}=12.5$ 得 $U=\dfrac{12.5-12}{1/\sqrt{100}}=5$

因为 $|U|=5>z_{\frac{\alpha}{2}}=1.645$.

所以 拒绝接受 H_0,即产品的平均重量有显著变化.

例 8.10 从一批灯泡中随机抽取 50 只,分别测量其寿命(单位:h) X,若 $X\sim N(\mu,\sigma^2)$,由抽出的样本算得样本平均值 $\bar{x}=1\ 900$,样本标准差 $s=490$.问能否认为这批灯泡的平均寿命为2 000h($\alpha=0.01$)?

分析 该问题为单正态总体(未知方差)均值的双侧假设检验.

解 $H_0:\mu=2\ 000$\ $H_1:\mu\neq 2\ 000$.

取检验统计量 $t=\dfrac{\bar{X}-2\ 000}{s/\sqrt{n}}$, $T\sim t(n-1)$.由 $\alpha=0.01$ 得 $\dfrac{\alpha}{2}=0.005$,结合自由度 $n-1=49$ 查 t 分布表得临界值 $t_{\frac{\alpha}{2}}(49)\approx t_{\frac{\alpha}{2}}(45)=2.689\ 6$.

因为 $\bar{x}=1\ 900$, $s=490$, $n=50$,算得

$$t=\frac{1\ 900-2000}{490/\sqrt{50}}=-1.443,$$

因为 $|t|=1.443<t_{\frac{\alpha}{2}}(45)=2.689\ 6$.

所以接受 H_0,即认为灯泡平均寿命为 2 000h.

例 8.11 甲、乙两机床加工同样的产品,两台机床加工的产品直径分别服从正态分布 $N(\mu_1,\sigma^2)$和 $N(\mu_2,\sigma^2)$,从两台机床加工的产品中分别抽取若干件,测得产品直径(单位:mm)如下.

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.

问甲、乙两机床加工的产品直径有无显著差异($\alpha=0.05$)?

分析 该问题为两正态总体(方差相等但未知)均值差异的双

侧检验问题.

解 $H_0: \mu_1 = \mu_2 \backslash\ H_1: \mu_1 \neq \mu_2$.

取检验统计量

$$t = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w} \quad \frac{n_1 n_2}{n_1 + n_2},$$

则 $t \sim t(n_1 + n_2 - 2)$.

由显著水平 $\alpha = 0.05$ **得** $\frac{\alpha}{2} = 0.025$,**结合自由度** $n_1 + n_2 - 2 = 8 + 7 - 2 = 13$,**查** t **分布表得临界值** $t_{0.025}(13) = 2.1604$.

计算得 $\overline{x} = 19.925$, $s_1 = 0.465$; $\overline{y} = 20$, $s_2 = 0.630$.

$$S_W = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{7 \times 0.465^2 + 6 \times 0.63^2}{13}$$

$$\approx 0.547,$$

$$\frac{n_1 n_2}{n_1 + n_2} = 1.932.$$

所以

$$t = \frac{\overline{x} - \overline{y}}{S_W} \quad \frac{n_1 n_2}{n_1 + n_2} = \frac{19.925 - 20}{0.547} \times 1.932$$

$$= -0.2649.$$

因为 $|t| = 0.2649 < t_{\frac{\alpha}{2}}(13) = 2.1604$.

所以接受 H_0,**即甲乙两机床加工的产品直径无显著差异.**

三、正态总体方差的假设检验

例 8.12 **某厂生产的铜丝,要求其折断力的方差不超过** 16,**今从某日生产的铜丝中随机抽取容量为** 9 **的一个样本,测得折断力**(**单位**: N)**数据为**: 289, 286, 285, 286, 284, 285, 286, 298, 292.**设总体服从正态分布,问该日生产的铜丝的折断力的方差是否合乎标准**($\alpha = 0.05$)?

分析 **这是单正态总体**(**未知期望** μ)**方差的单侧检验问题.**

解　$H_0: \sigma^2 \le 16$，$H_1: \sigma^2 > 16$.

取检验统计量 $\chi^2 = \dfrac{(n-1)s^2}{16}$，则 χ^2 服从 $\chi^2(n-1)$.由显著水平 $\alpha = 0.05$，结合自由度 $n-1=8$ 查 χ^2 分布表得临界值 $\chi^2_{0.05}(8) = 15.507$.

计算得 $\bar{x} = 288.11$，　$s^2 = 20.3611$.

所以
$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{8\times 20.361\,1}{16} = 10.180\,6.$$

因为
$$\chi^2 = 10.180\,6 < \chi^2_\alpha(8) = 15.507.$$

所以接受 H_0，即认为铜丝折断力的方差不超过 16.

例 8.13　对两批同类电子元件电阻进行测试，各抽 6 件，测得电阻值(单位:Ω)如下.

A 批:0.140,0.138,0.143,0.141,0.144,0.137;

B 批:0.135,0.140,0.142,0.136,0.138,0.141.

已知元件的电阻服从正态分布.设 $\alpha = 0.05$，问:

(1) 两批电子元件电阻的方差是否相等?

(2) 两批元件的平均电阻是否有显著差异?

分析　这里有两个问题，一是两正态总体方差(未知期望)相等的双侧检验问题.二是两正态总体(未知方差，但由前假设检验可知方差相等)均值差异的双侧检验问题.

解　(1) $H_0: \sigma_1^2 = \sigma_2^2$，$H_1: \sigma_1^2 \ne \sigma_2^2$.

取检验统计量 $F = \dfrac{s_大^2}{s_小^2}$，由 F 服从 $F(n_大 - 1, n_小 - 1)$.再由 $\alpha = 0.05$ 得 $\dfrac{\alpha}{2} = 0.025$，结合自由度 $(n_大 - 1, n_小 - 1)$ 可得临界值 $F_{\frac{\alpha}{2}}(n_大 - 1, n_小 - 1)$.

算得
$$\bar{x} = 0.140\,5,\quad s_1 = 0.002\,738;$$
$$\bar{y} = 0.138\,7,\quad s_2 = 0.002\,805.$$

得
$$F = \frac{s_大^2}{s_小^2} = \frac{s_2^2}{s_1^2} = \frac{0.002\,805^2}{0.002\,738^2} \approx 1.049\,5.$$

由$\frac{\alpha}{2}=0.025$，结合自由度$(n_2-1, n_1-1)=(5,5)$查 F 分布表得上侧临界值

$$F_{\frac{\alpha}{2}}(5,5)=7.15.$$

因为 $F=1.0495<F_{\frac{\alpha}{2}}(5,5)=7.15$.

所以接受 H_0，即认为两批元件电阻的方差相等.

(2) $H_0: \mu_1=\mu_2$ \ $H_1: \mu_1\neq\mu_2$.

取检验统计量 $t=\frac{\overline{x}-\overline{y}}{S_W}\sqrt{\frac{n_1 n_2}{n_1+n_2}}$，则 $t\sim t(n_1+n_2-2)$.由显著水平 $\alpha=0.05$ 得$\frac{\alpha}{2}=0.025$，结合自由度 $n_1+n_2-2=8$ 查 t 分布表得临界值

$$t_{\frac{\alpha}{2}}(8)=2.3060.$$

算得

$$\begin{aligned}S_W&=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\\&=\sqrt{\frac{4\times0.002738^2+4\times0.002805^2}{8}}\\&=0.002772.\end{aligned}$$

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}=\sqrt{\frac{6\times6}{6+6}}\approx1.732.$$

所以

$$\begin{aligned}t&=\frac{\overline{x}-\overline{y}}{S_W}\cdot\sqrt{\frac{n_1 n_2}{n_1+n_2}}=\frac{0.1405-0.1387}{0.002772}\times1.732\\&=1.125.\end{aligned}$$

因为 $|t|=1.125<t_{\frac{\alpha}{2}}(8)=2.3060$.

所以接受 H_0，即两批元件的平均电阻无显著差异.

注　第一问题(方差相等检验问题)是双侧检验，确定临界值时，原应该根据$\frac{\alpha}{2}$以及自由度(n_1-1, n_2-1)查 F 分布表得上、下

临界值 $F_{\frac{\alpha}{2}}(n_1-1, n_2-1)$和 $F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)$.但由于选取的统计量是 $F=\frac{S_{大}^2}{S_{小}^2}$,所以查 F 分布表时,只须确定上侧临界值 $F_{\frac{\alpha}{2}}(n_{大}-1, n_{小}-1)$.因为此时

$$F=\frac{S_{大}^2}{S_{小}^2}<1<F_{1-\frac{\alpha}{2}}(n_{大}-1, n_{小}-1)$$

肯定不发生.这样可简化检验过程.

小结　为便于理解并记住假设检验问题解题过程,可将之归纳为如下三个步骤进行,这样使得解题思路清晰,表述简洁而层次分明.

(1) 提出假设

根据问题的要求提出相应的假设.

(2) 推理分析

在假设成立的前提下,根据问题的类型选取合适的检验统计量.然后将显著水平化成查表所需形式,再结合自由度查相应的分布表得接受域与拒绝域的分界点(临界值).

(3) 检验判断

由已知样本值,计算选取的检验统计量值.

若此值落入接受域,则接受 H_0.

若此值落入拒绝域,则拒绝接受 H_0.

四、总体分布假设的 χ^2 检验法

例 8.14　在一小型电话交换台的呼唤次数按每分钟统计如下表:

每分钟呼唤次数	0	1	2	3	4	5	6
频　　数	8	16	17	10	6	2	1

用 χ^2 检验法检验每分钟内的电话呼唤次数服从泊松分布的假设 $(\alpha=0.05)$.

解　$H_0: X \sim p(\lambda)$.

因为 $\mu = E(X) = \lambda$,而 μ的极大似然估计值为

$$\hat{\lambda} = \hat{\mu} = \bar{x}$$

$$= \frac{0\times 8 + 1\times 16 + 2\times 17 + 3\times 10 + 4\times 6 + 5\times 2 + 6\times 1}{8 + 16 + 17 + 10 + 6 + 2 + 1}$$

$$= 2 .$$

因为 $X = 5$, $X = 6$ 的频数 $m_i < 5$,将它们就近并入 $X = 4$ 而成为 $X \geq 4$ 组 .所以 $r = 5$, $l = 1$.由显著水平 $\alpha = 0.05$ 和自由度 $r - l - 1 = 5 - 1 - 1 = 3$,查 χ^2 分布表得临界值 $\chi^2_\alpha(3) = 7.815$.

查泊松分布表得:

每分钟呼唤次数 X	0	1	2	3	≥ 4
实际频数 m_i	8	16	17	10	9
理论频率 p_i	0.135 3	0.270 7	0.270 7	0.180 5	0.142 9
理论频数 np	8.118	16.242	16.242	10.830	8.574

$$\chi^2 = \frac{(8 - 8.118)^2}{8.118} + \frac{(16 - 16.242)^2}{16.242} + \frac{(17 - 16.242)^2}{16.242} + \frac{(10 - 10.824)^2}{10.824} + \frac{(9 - 8.574)^2}{8.574} = 0.124\,6 .$$

因为
$$\chi^2 = 0.1246 < \chi^2_\alpha(3) = 7.815 .$$

所以接受 H_0,即可认为每分钟内电话呼唤次数服从泊松分布 .

习 题 选 解

(**8 - 1**)　一种燃料的辛烷等级服从正态分布,其平均等级为 98.0,标准差为 0.8 .今从一批新油中抽检 25 桶,算得样本均值为 97.7 .假定标准差与原来一样,问新油的辛烷平均等级是否比原燃料平均等级偏低($\alpha = 0.05$) .

分析　因样本标准差与原来一样,可视为方差已知 .故该问题

为单正态总体(已知方差)均值的单侧检验问题.

解 $H_0:\mu\geq 98$\ $H_1:\mu<98$.

取检验统计量 $V=\dfrac{\overline{x}-98}{0.8/\sqrt{25}}$,则 $V\sim N(0,1)$由显著水平 $\alpha=0.05$ 得 $1-\dfrac{\alpha}{2}=0.975$,查标准正态分布表得临界值 $z_{\frac{\alpha}{2}}=1.96$.

$$U=\frac{\overline{x}-98}{0.8/\sqrt{25}}=\frac{97.7-98}{0.8/5}=-1.875.$$

因为 $$U=-1.875>-z_{\frac{\alpha}{2}}=-1.96.$$

所以接受 H_0,即不能认为新油的辛烷平均等级比原燃料平均等级偏低.

(8-2) 设有来自正态总体 $X\sim N(\mu,\sigma^2)$,容量为 100 的样本,样本均值 $\overline{x}=2.7$, μ、σ^2 均未知,而 $\sum\limits_{i=1}^{100}(x_i-\overline{x})^2=225$.在 $\alpha=0.05$ 下,检验下列假设.

(1) $H_0:\mu=3$\ $H_1:\mu\neq 3$.

(2) $H_0:\sigma^2=2.5$\ $H_1:\sigma^2\neq 2.5$.

分析 (1) 为单正态总体(未知方差)期望 μ 的双侧假设检验.(2) 为单正态总体(未知期望)方差 σ^2 的双侧假设检验.

解 (1) $H_0:\mu=3$\ $H_1:\mu\neq 3$

选取检验统计量 $t=\dfrac{\overline{X}-3}{s/\sqrt{100}}$,则 $t\sim t(100-1)$.由 $\alpha=0.05$ 得 $\dfrac{\alpha}{2}=0.025$,结合自由度 $n-1=99$,查 t 分布表得 $t_{\frac{\alpha}{2}}(99)\approx t_{\frac{\alpha}{2}}(100)=1.984$.

由 $n=100$, $\overline{x}=2.7$, $\sum\limits_{i=1}^{100}(x_i-\overline{x})=225$ 得

$$s=\sqrt{\frac{1}{99}\times 225}=1.508.$$

所以 $$t=\frac{\overline{x}-\mu_0}{s/\sqrt{n}}=\frac{2.7-3}{1.508/\sqrt{100}}=-1.9894.$$

而 $$|t| = 1.9894 > t_{\frac{\alpha}{2}}(99) = 1.984 .$$

所以拒绝 H_0,即认为 $\mu \neq 3$.

(2) $H_0: \sigma^2 = 2.5 \backslash\ \sigma^2 \neq 2.5$.

取统计量 $\chi^2 = \dfrac{(n-1)s^2}{2.5}$,则 $\chi^2 \sim \chi^2(n-1)$由 $\alpha = 0.05$ 得 $\dfrac{\alpha}{2} = 0.025, 1 - \dfrac{\alpha}{2} = 0.975$,结合自由度 $n - 1 = 99$,查 χ^2 分布表得临界值

$$\chi^2_{0.025}(99) \approx \chi^2_{0.025}(100) = 129.56 .$$

$$\chi^2_{0.975}(99) \approx \chi^2_{0.975}(100) = 74.22 .$$

$$\chi^2 = \frac{(n-1)s^2}{2.5} = \frac{\sum_{i=1}^{100}(x_i - \overline{x})^2}{2.5} = \frac{225}{2.5} = 90 .$$

因为 $$\chi^2_{0.975}(99) = 74.22 < \chi^2 = 90 < \chi^2_{0.975}(99) = 129.56 .$$

所以接受 H_0,即认为方差 $\sigma^2 = 2.5$.

(8-3) 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随机抽 9 根,测得样本标准差 $s = 0.007\Omega$.设总体服从正态分布,能认为这批导线的标准差显著偏大吗($\alpha = 0.05$)?

分析 这是单正态总体(未知期望)方差的单侧检验问题 .

解 $H_0: \sigma \leq 0.005 \backslash\ H_1: \sigma > 0.005$.

取检验统计量 $\chi^2 = \dfrac{(n-1)s^2}{0.005^2}$,则 $\chi^2 \sim \chi^2(8)$.由显著水平 $\alpha = 0.05$和自由度 $n - 1 = 8$ 查 χ^2 分布表得临界值

$$\chi^2_{0.05}(8) = 15.507 .$$

$$\chi^2 = \frac{(n-1)s^2}{0.005^2} = \frac{8 \times 0.007^2}{0.005^2} = 15.68,$$

因为 $$\chi^2 = 15.68 > \chi^2_{0.05}(8) = 15.507 .$$

所以拒绝 H_0,即认为这批导线的标准差显著偏大 .

(8-4) 机器包装食盐,假设每袋盐净重(单位:g)服从正态分布,规定每袋标准含量为 500 g,标准差不得超过 10 g,某天开工

后，随机抽取 9 袋，测得净重(单位：g)如下：497，507，510，475，515，484，488，524，491．检验假设($\alpha=0.05$)．

(1) $H_0:\mu=500$\ $H_1:\mu\neq500$；

(2) $H_0:\sigma\leq10$\ $H_1:\sigma>10$．

以判断这天包装机工作是否正常(假定(1)、(2)有一项拒绝 H_0，即认为不正常)．

分析 (1) 为单正态总体(未知方差)均值的双侧检验问题，(2) 为单正态总体(未知期望)方差的单侧检验问题．

解 (1) $H_0:\mu=500$\ $H_1:\mu\neq500$．

取检验统计量 $t=\dfrac{\overline{X}-500}{s/\sqrt{n}}$，则 $t\sim t(8)$．由 $\alpha=0.05$ 得 $\dfrac{\alpha}{2}=0.025$，结合自由度 $n-1=8$ 查 t 分布表得临界值

$$t_{\frac{\alpha}{2}}(8)=2.306.$$

算得 $\overline{x}=499$，$s=16.031$，

所以
$$t=\frac{\overline{x}-\mu_0}{s/\sqrt{n}}=\frac{499-500}{16.031/\sqrt{9}}\approx-0.1871.$$

因为
$$|t|=0.1871<t_{\frac{\alpha}{2}}(8)=2.306.$$

所以接受 H_0，即可认为当天每袋食盐平均重量为 500 g．

(2) $H_0:\sigma^2\leq100$\ $H_1:\sigma^2>100$．

取统计量 $\chi^2=\dfrac{(n-1)s^2}{\sigma_0^2}=\dfrac{(n-1)s^2}{100}$，则 $\chi^2\sim\chi^2(8)$，由 $\alpha=0.05$ 结合自由度 $n-1=8$ 查 χ^2 分布表得临界值

$$\chi^2_{0.05}(8)=15.507.$$

$$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}=\frac{8\times16.031^2}{100}=20.58$$

因为
$$\chi^2=20.58>\chi^2_{0.05}(8)=15.507.$$

所以拒绝 H_0，即认为当天标准差超过 10 g．

综合(1)、(2)可知该天包装机工作不正常．

(8-5) 改进某种金属的热处理方法，要检验抗拉强度(单

位:Pa)有无显著提高.在改进前后各取12个样品,测量并计算结果如下.

改进前: $\overline{x} = 28.67, \sum_{i=1}^{12} (x_i - \overline{x})^2 = 66.64$.

改进后: $\overline{y} = 31.75, \sum_{i=1}^{12} (y_i - \overline{y})^2 = 112.25$.

假定热处理前金属抗拉强度都服从正态分布.取 $\alpha = 0.05$.问:

(1) 处理前后方差是否相等?

(2) 处理后抗拉强度有无显著提高?

分析 (1) 为两正态总体方差(未知期望)差异的双侧检验问题.(2) 为两正态总体期望(未知方差但方差相等)差异的单侧检验问题.

解 (1) $H_0: \sigma_1^2 = \sigma_2^2 \backslash \; H_1: \sigma_1^2 \neq \sigma_2^2$.

取检验统计量

$$F = \frac{s_{大}^2/\sigma^2}{s_{小}^2/\sigma^2} = \frac{s_{大}^2}{s_{小}^2}, \text{则 } F \sim F(n_{大} - 1, n_{小} - 1).$$

$$s_1^2 = \frac{66.64}{11} \approx 6.058 = s_{小}^2.$$

$$s_2^2 = \frac{112.25}{11} \approx 10.205 = s_{大}^2.$$

所以
$$F = \frac{s_{大}^2}{s_{小}^2} = \frac{s_2^2}{s_1^2} \approx 1.684.$$

由 $\alpha = 0.05$ 得 $\frac{\alpha}{2} = 0.025$,结合两自由度(11,11),查 F 分布表得临界值

$$F_{\frac{\alpha}{2}}(11,11) \approx 3.48.$$

因为
$$F = 1.684 < F_{\frac{\alpha}{2}}(11,11) = 3.48.$$

所以接受 H_0,即认为处理前后总体方差相等.

(2) $H_0: \mu_1 \geq \mu_2 \backslash \; H_1: \mu_1 < \mu_2$.

取检验统计量 $t = \frac{\overline{X} - \overline{Y}}{S_W} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$,则 $t \sim t(22)$.由显著水平

$\alpha = 0.05$，结合自由度 $n_1 + n_2 - 2 = 22$ 查 t 分布表得临界值

$$t_\alpha(22) = 1.7171.$$

$$S_W = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{66.64 + 112.25}{22}}$$

$$\approx 2.8516.$$

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \sqrt{\frac{12 \times 12}{12 + 12}} = 2.4495.$$

所以 $$t = \frac{28.67 - 31.75}{2.8516} \times 2.4495 \approx -2.6457.$$

因为 $t = -2.6457 < -t_\alpha(22) = -1.7171$.

所以拒绝 H_0，即处理后抗检验强度有显著提高.

(8－6) 某船厂根据以前的显示资料知道本厂生产的农用船只分别以 20%、28%、8%、12% 和 32% 的比例卖给 A、B、C、D、E 五个地区，从今年生产的农用船中，观察到其中 500 艘分别售于上述五个地区 120、123、43、66、148 艘.问今年这五个地区的销售比例是否与以往有显著不同($\alpha = 0.05$)?

分析 销售比例实际为概率分布，所以这是总体分布的假设检验问题.

解 H_0:

地区	A	B	C	D	E
概率 p_i	0.2	0.28	0.08	0.12	0.32

频数分布如下表($n = 500$).

地区	A	B	C	D	E
理论频数 np_i	100	140	40	60	160
实际频数 m_i	120	123	43	66	148

取检验统计量 $\chi^2=\sum_{i=1}^{5}\frac{(m_i-np_i)^2}{np_i}$.则

$$\chi^2\sim\chi^2(r-l-1).\text{其中 } r=5, l=0.$$

由显著水平 $\alpha=0.05$ 和自由度 $r-l-1=4$,查 χ^2 分布表得临界值 $\chi^2_\alpha(4)=9.488$.

$$\chi^2=\frac{(120-100)^2}{100}+\frac{(123-140)^2}{140}+\frac{(43-40)^2}{40}+\frac{(66-60)^2}{60}+\frac{(148-160)^2}{160}\approx 7.789.$$

因为
$$\chi^2=7.789<\chi^2_\alpha(4)=9.488.$$

所以接受 H_0,即今年这五个地区的销售比例与以往无显著不同.

(8-7) 在孟德尔的豌豆试验中,他曾对 10 株豌豆统计了黄色豌豆和青色豌豆的频数,黄色豌豆 355 颗,青色豌豆 123 颗,总计 478 颗.根据孟德尔的理论,黄∶青 = 3∶1.试检验这一假设($\alpha=0.05$).

分析 由黄、青豌豆之比可得黄、青豌豆理论上的概率分布 $p(\text{黄})=\frac{3}{4}$, $p(\text{青})=\frac{1}{4}$.这是个总体分布的假设检验问题.

解 $H_0: p(\text{黄})=\frac{3}{4}, p(\text{青})=\frac{1}{4}$.

依题意可得理论与实际频数如下表.

豌豆	黄	青
理论频数 np_i	$478\times\frac{3}{4}$	$478\times\frac{1}{4}$
实际频数 m_i	355	123

取检验统计量 $\chi^2=\sum_{i=1}^{r}\frac{(m_i-np_i)^2}{np_i}$, $\chi^2\sim\chi^2(r-l-1)$.

现 $r=2, l=0$.由 $\alpha=0.05$ 和自由度 $r-l-1=1$,查 χ^2 分布表得

临界值 $\chi^2_\alpha(1)=3.841$.

计算得 $$\chi^2=\frac{\left(355-478\times\frac{3}{4}\right)^2}{478\times\frac{3}{4}}+\frac{\left(123-478\times\frac{1}{4}\right)^2}{478\times\frac{1}{4}}$$

$$\approx 0.1367.$$

因为 $\chi^2=0.1367<\chi^2_{0.05}(1)=3.841$.

组　　别	频数
10.93～10.95	5
10.95～10.97	8
10.97～10.99	20
10.99～11.01	34
11.01～11.03	17
11.03～11.05	6
11.05～11.07	6
11.07～11.09	4

所以接受 H_0,即认为黄∶青 $=3∶1$.

(8-8) 对某汽车零件制造厂生产的汽缸螺检口径进行抽样检验,测得 100 个数据分组列表如右:试检验螺检口径是否服从正态分布($\alpha=0.05$)?

分析 这是总体分布检验问题.

解 H_0:螺检口径 $X\sim N(\mu,\sigma^2)$.

对参数 μ、σ^2 用极大似然估计得:

$$\mu=\overline{x}\approx 11,\quad \sigma^2=\frac{1}{n}\sum(x_i-\overline{x})^2\approx 0.03^2.$$

取检验统计量

$$\chi^2=\sum_{i=1}^{r}\frac{(m_i-np_i)^2}{np_i},\text{则 } \chi^2\sim\chi^2(r-l-1).$$

由样本数据统计计算如下表:

$$p_1=\Phi\left(\frac{10.95-11}{0.03}\right)-\Phi(-\infty)$$

$$p_i=\Phi\left(\frac{b_i-11}{0.03}\right)-\Phi\left(\frac{a_i-11}{0.03}\right),\ i=2,3,\cdots,6.$$

$$p_7=\Phi(+\infty)-\Phi\left(\frac{11.05-11}{0.03}\right)$$

查标准正态表可得 p_i.

组　　别	实际频数 m_i	理论频数 p_i	理论频数 np_i	$(m_i-np_i)^2/np_i$

$-\infty$～10.95	5	0.047 5	4.75	0.013
10.95～10.97	8	0.111 2	11.12	0.875
10.97～10.99	20	0.212 0	21.20	0.068
10.99～11.01	34	0.258 6	25.86	2.562
11.01～11.03	17	0.212 0	21.20	0.832
11.03～11.05	6	0.111 2	11.12	2.357
11.05～11.07 11.07～$+\infty$	6 4 10	0.047 5	4.75	5.803

$$\chi^2 = 0.013 + 0.875 + 0.068 + 2.562 + 0.832 + 2.357 + 5.803 = 12.51 .$$

由 $\alpha = 0.05$ 和自由度 $r - l - 1 = 7 - 2 - 1 = 4$,查 χ^2 分布表得临界值 $\chi^2_{0.05}(4) = 9.488$.

因为
$$\chi^2 = 12.51 > \chi^2_{0.05}(4) = 9.488 .$$
所以拒绝 H_0,即认为 X 不服从正态分布 .

注　此题若将显著水平改为 $\alpha = 0.01$,结合自由度 $r - l - 1 = 4$,查 χ^2 分布表得 $\chi^2_{0.01}(4) = 13.277$.

因为
$$\chi^2 = 12.51 < \chi^2_{0.01}(4) = 13.277 .$$
所以接受 H_0,即认为 X 服从正态分布 .

试 题 选 解

[8－1]　填空题

1 .设总体 X 服从正态分布 $N(\mu, \sigma^2)$,其中 μ 未知, $X_1, X_2, \cdots, X_n$ 为其样本 .若假设检验问题为 $H_0: \sigma^2 = 1 \backslash\ H_1: \sigma^2 \neq 1$.则采用的检验统计量应为(　　) .

答　$\chi^2 = (n-1)s^2$.

分析　这是单正态总体(未知期望 μ)方差的双侧检验,应选取检验统计量 .

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = (n-1)s^2 \left(\text{或} = \sum_{i=1}^{n}(x_i - \overline{x})^2\right) .$$

2．设某个假设检验问题的拒绝域为 W，且当原假设 H_0 成立时，样本值 $x_1, x_2, \cdots, x_n$ 落入 W 的概率为 0.15，则犯第一类错误的概率为(　　)．

答　0.15．

分析　第一类错误的概率

$$\alpha = P\{\text{样本统计量} \in W \mid H_0 \text{成立}\} = 0.15 .$$

3．设样本 $X_1, X_2, \cdots, X_n$ 来自正态总体 $N(\mu, 1)$，假设检验问题为：

$$H_0: \mu = 0 \qquad H_1: \mu \neq 0 .$$

则在 H_0 成立的条件下，对显著水平 α，拒绝域 W 应为(　　)．

答　$\{|u| > z_{\frac{\alpha}{2}}\}$．

分析　这是单正态总体(已知方差 $\sigma^2 = 1$)期望 μ 的双侧检验问题．应选取统计量

$$U = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \overline{x}\sqrt{n}, \qquad U \sim N(0,1) .$$

由 α 得 $1 - \frac{\alpha}{2}$ 值，查标准正态分布表得临界值 $z_{\frac{\alpha}{2}}$ 使

$$P\{|u| > z_{\frac{\alpha}{2}}\} = \alpha .$$

所以拒绝域 W 为$\{|u| > z_{\frac{\alpha}{2}}\}$．其中 $u = \overline{x}\sqrt{n}$．

[8-2]　计算题

1．实验室需要购置一台精度为 10^{-1}g 的天平，为了使所购天平达到所需要求，在选购时，将一重为 μg(μ 未知)的物体在一台天平上独立重复地称 10 次，这 10 次称量结果的样本方差 $s^2 = 1.6 \times 10^{-3}$．假定天平称量结果服从正态分布 $N(\mu, \sigma^2)$．

(1) 这台天平的精度是否达到了实验室所需要求($\alpha = 0.01$)？

(2) 给出 σ^2 的置信水平为 99% 的单侧置信下限(已知 $\chi^2_{0.01}(9) = 2.088$)．

分析　这一道题既含假设检验问题，又含参数区间估计问题，

自考中这种题型较多见，首先要能区分谁是假设检验，谁是参数估计？本题中：(1) 为单正态总体（未知期望）方差的单侧假设检验问题．(2) 为单正态总体（未知期望）方差的单侧置信下限问题（区间估计）．

解 (1) $H_0:\sigma^2\geq 10^{-2}\backslash\quad H_1:\sigma^2<10^{-2}$．

取检验统计量 $\chi^2=\frac{(n-1)s^2}{10^{-2}}$，则 $\chi^2\sim\chi^2(9)$．由 $\alpha=0.01$ 以及自由度 $n-1=9$ 查 χ^2 分布表得临界值 $\chi^2_{0.01}(9)=2.088$．拒绝域 W 为 $\chi^2<\chi^2_{0.01}(9)$．

$$\chi^2=\frac{(n-1)s^2}{10^{-2}}=\frac{9\times 1.6\times 10^{-3}}{10^{-2}}=1.44.$$

因为
$$\chi^2=1.44<\chi^2_{0.01}(9)=2.088.$$

所以拒绝 H_0，即该天平的精度达到了所需要求．

(2) 取统计量 $\chi^2=\frac{(n-1)s^2}{\sigma^2}$，则 $\chi^2\sim\chi^2(9)$．

由 $1-\alpha=0.99$ 得 $\alpha=0.01$，结合自由度 $n-1=9$ 查 χ^2 分布表得上侧临界值 $\chi^2_{0.01}(9)=2.088$．使

$$P\ \frac{(n-1)s^2}{\sigma^2}\leq\chi^2_{0.01}(9)\ =0.99.$$

即
$$P\ \sigma^2\geq\frac{(n-1)s^2}{\chi^2_{0.01}(9)}\ =0.99.$$

因为
$$\frac{(n-1)s^2}{\chi^2_{0.01}(9)}=\frac{9\times 1.6\times 10^{-3}}{2.088}=6.9\times 10^{-3}.$$

所以 σ^2 的置信度为 99％的单侧置信下限为 6.9×10^{-3}．

2．用传统工艺加工的某种水果罐头中，每瓶平均 V_C（维生素 C）含量为 19 mg，现改进了加工工艺，抽查了 16 瓶罐头，得

$$\bar{x}=20.8,\quad s^2=1.6^2.$$

已知水果罐头中的 V_C 含量服从方差未知的正态分布．

(1) 问新工艺下 V_C 含量是否有明显提高（$\alpha=0.005$）？

(2) 给出新工艺下平均 V_C 含量的置信水平为 0.99 的双侧置

信区间(已知 $t_{0.005}(15)=2.95$).

分析 (1) 为单正态总体(未知方差)期望 μ 的单侧假设检验问题.(2) 为单正态总体(未知方差)期望 μ 的双侧置信区间问题.

解 (1) $H_0: \mu \leq 19$\ $H_1: \mu > 19$.

取检验统计量 $t=\dfrac{\overline{X}-19}{s/\sqrt{n}}$,则 $t \sim t(n-1)$,由显著水平 $\alpha=0.005$和自由度 $n-1=15$,查 t 分布表得临界值

$$t_{0.005}(15)=2.95 .\text{拒绝域 } W \text{ 为 } t > t_{0.005}(15).$$

$$t=\frac{\overline{x}-19}{s/\sqrt{n}}=\frac{20.8-19}{1.6/\sqrt{16}}=4.5$$

因为 $$t=4.5>t_{0.005}(15)=2.95.$$

所以拒绝 H_0,即认为新工艺下 V_C 含量有明显提高.

(2) 取统计量 $t=\dfrac{\overline{X}-\mu}{s/\sqrt{n}}$,则 $t \sim t(n-1)$.

由 $$1-\alpha=0.99 \quad 得 \quad \frac{\alpha}{2}=0.005,$$

结合自由度 $n-1=15$,查 t 分布表得临界值 $t_{0.005}(15)=2.95$.

因为 $\overline{x}=20.8$, $s=1.6$.

置信上限:

$$\hat{\mu}_2=\overline{x}+\frac{s}{\sqrt{n}}t_{0.005}(15)=20.8+\frac{1.6}{\sqrt{16}}\times 2.95$$
$$=20.8+1.18=21.98.$$

置信下限:

$$\hat{\mu}_1=\overline{x}-\frac{s}{\sqrt{n}}t_{0.005}(15)=20.8-1.18=19.62.$$

所以 V_C 含量置信度为 0.99 的置信区间为[19.62,21.98].

3. 一种元件,要求其寿命不得低于 1 000 小时.现从一批这种元件中随机抽取 25 件,测得其寿命平均值为 950 小时.已知该种元件的寿命服从方差为 $\sigma^2=100^2$ 的正态分布.

(1) 试在显著水平 $\alpha=0.05$ 下,确定这批元件是否合格?

(2) 给出元件平均寿命的置信水平为 0.95 的双侧置信区间(已知 $u_{0.95}=1.65$, $u_{0.975}=1.96$).

分析 (1) 为单正态总体(已知方差)期望 μ 的单侧假设检验问题.(2) 为单正态总体(已知方差)期望 μ 的双侧置信区间问题.

解 (1) $H_0: \mu \geq 1\,000 \backslash \quad H_1: \mu < 1\,000$.

取检验统计量 $u=\dfrac{\overline{x}-1\,000}{\sigma/\sqrt{n}}=\dfrac{\overline{x}-1\,000}{100/\sqrt{25}}$,则 $U \sim N(0,1)$.

由显著水平 $\alpha=0.05$ 得 $1-\alpha=0.95$,查标准正态分布表得临界值 $u_{0.95}=1.65$,拒绝域为 $W=\{u<-u_{0.95}\}$.

已知 $n=25$, $\overline{x}=950$,算得

$$u=\frac{\overline{x}-1000}{\sigma/\sqrt{n}}=\frac{950-1000}{100/\sqrt{25}}=-2.5.$$

因为 $$u=-2.5<-u_{0.95}=-1.65.$$

所以拒绝 H_0,即这批元件寿命低于 1 000 小时,所以这批元件不合格.

(2) 取统计量 $u=\dfrac{\overline{x}-\mu}{\sigma/\sqrt{n}}=\dfrac{\overline{x}-\mu}{100/\sqrt{25}}$,则 $u \sim N(0,1)$ 由 $1-\alpha=0.95$ 得 $1-\dfrac{\alpha}{2}=0.975$,查标准正态分布表得临界值 $u_{0.975}=1.96$.

已知 $\overline{x}=950$, $n=25$, $\sigma=100$.

所以置信上限:

$$\hat{\mu}_2=\overline{x}+\frac{\sigma}{\sqrt{n}}u_{0.975}=950+\frac{100}{\sqrt{25}}\times 1.96$$
$$=950+39.2=989.2.$$

置信下限:

$$\hat{\mu}_1=\overline{x}-\frac{\sigma}{\sqrt{n}}u_{0.975}=950-39.2=910.8.$$

所以元件平均寿命的置信水平为 0.95 的置信区为[910.8, 989.2].

4. 某钢厂为了确定一种新工艺能否提高钢的得率,在同一

平炉上用新老工艺交替进行了 10 次试验,其得率如下.

老工艺:78.1,72.4,70.2,74.3,77.4,

78.4,76.0,75.5,76.7,77.3;

新工艺:79.1,81.0,77.3,79.1,80.0,

79.1,79.1,77.3,80.2,82.1.

设这两个样本独立,且分别来自正态总体 $X \sim N(\mu_1, \sigma^2)$和 $Y \sim N(\mu_2, \sigma^2)$ (μ_1、μ_2、σ^2 均未知),经计算得 $\bar{x} = 76.23$, $s_1^2 = 3.325$, $\bar{y} = 79.43$, $s_2^2 = 2.225$.问新工艺是否能显著提高钢的得率($\alpha = 0.05$)($t_{0.05}(18) = 1.7341$)?

分析　这是两正态总体(未知方差但方差相等)期望差的单侧假设检验问题.

解　$H_0: \mu_1 \geq \mu_2$ \ $H_1: \mu_1 < \mu_2$.

取检验统计量 $t = \dfrac{\bar{x} - \bar{y} - 0}{S_W} \quad \dfrac{n_1 n_2}{n_1 + n_2}$,则 $t \sim t(n_1 + n_2 - 2)$.由 $\alpha = 0.05$ 和自由度 $n_1 + n_2 - 2 = 18$ 查 t 分布表得 $t_{0.05}(18) = 1.7341$.拒接域为 $W = \{t < -t_{0.05}(18)\}$.

算得

$$\bar{x} = 76.23, \quad s_1^2 = 3.325, \quad \bar{y} = 79.43, \quad s_2^2 = 2.225.$$

$$S_W = \quad \frac{(10-1)s_1^2 + (10-1)s_2^2}{10+10-2}$$

$$= \quad \frac{3.325 + 2.225}{2} \approx 1.666.$$

$$\frac{n_1 n_2}{n_1 + n_2} = \quad \frac{10 \times 10}{10+10} = 2.236.$$

所以 $t = \dfrac{\bar{x} - \bar{y}}{S_W} \quad \dfrac{n_1 n_2}{n_1 + n_2} = \dfrac{76.23 - 79.43}{1.666} \times 2.236 = -4.295$.

因为 $t = -4.295 < -t_{0.05}(18) = -1.7341$.

所以拒绝 H_0,接受 $H_1: \mu_1 < \mu_2$,即认为新工艺提高了钢的得率.

第九章　回归分析与方差分析

内 容 提 要

一、知识脉络

- 一元线性回归
  - 一元线性回归模型
  - 回归参数 a,b 的最小二乘估计
  - 相关系数的计算
  - 回归方程的显著性检验
  - 预测
- 单因素方差分析
  - 单因素方差分析模型
  - 单因素方差分析方法
  - 单因素方差分析中的参数估计

二、知识要点

(一) 一元线性回归模型

设 x 与 y 是两个有相关关系的变量,其中 x 为非随机变量, y 为随机变量, y 与 x 的关系是:

$$y = a + bx + \varepsilon .$$

其中 a, b 是未知常数, ε 是服从正态分布 $N(0, \sigma^2)$ 的随机变量, x 与 y 的这种统计关系称为一元线性回归 .

(二) 参数 a, b 的最小二乘估计

设 n 个点 $(x_i, y_i)(i = 1, 2, \cdots, n)$ 为已知,记

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad l_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2,$$

$$l_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2, \quad l_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 .$$

则 $\hat{b} = \frac{l_{xy}}{l_{xx}}$, $\hat{a} = \overline{y} - \hat{b}\,\overline{x}$,经验线性回归:

$$\hat{y} = \hat{a} + \hat{b}x .$$

(三) 样本相关系数 r

$$r = \frac{l_{xy}}{l_{xx} \quad l_{xy}}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2 \quad \sum_{i=1}^{n}(y_i - \overline{y})^2}, \quad |r| \leq 1 .$$

$|r|$ 越接近 1,说明线性回归效果越好;

$|r|$ 越接近 0,表明线性回归效果越差 .

(四) 回归方程的显著性检验

记 $\quad Q = l_{yy}(1 - r^2), \quad U = l_{yy} r^2, \quad s = \frac{Q}{n-2} .$

1 . F 检验法

$$H_0: b=0, \quad F=(n-2)\frac{U}{Q} \sim F(1, n-2).$$

对于显著性水平 α，当 $F > F_\alpha(1, n-2)$ 时，拒绝 H_0，即认为回归方程显著；否则，接受 H_0，即认为回归方程不显著.

2. t 检验法

在 H_0 成立时

$$t=\frac{\hat{b}}{s} \quad l_{xx} \sim t(n-2).$$

对于显著性水平 α，当 $|t| > t_{\frac{\alpha}{2}}(n-2)$ 时，拒绝 H_0，即认为回归方程显著；当 $|t| \leq t_{\frac{\alpha}{2}}(n-2)$ 时，接受 H_0，即认为回归方程不显著.

(五) 单因素方差分析模型

设因素 A 有 p 个不同的水平 $A_1, A_2, \cdots, A_p$，x_{ij} 表示 A_j 下的第 i 个样本，$x_{ij}(i=1,2,\cdots,r; j=1,2,\cdots,p)$ 相互独立. 所讨论的单因素分析模型为

$$\begin{aligned} x_{ij} &= \mu_j + \varepsilon_{ij} \qquad & i=1,2,\cdots,r \\ \varepsilon_{ij} &\sim N(0, \sigma^2) \qquad & j=1,2,\cdots,p \end{aligned}$$

其中 μ_j 及 σ^2 均为未知常数，且各 ε_{ij} 相互独立. 假设检验

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_p.$$

(六) 单因素方差分析方法

记 $\quad \bar{x}_j = \frac{1}{r}\sum_{i=1}^{r} x_{ij}, \quad \bar{x} = \frac{1}{rp}\sum_{j=1}^{p}\sum_{i=1}^{r} x_{ij}$

$$S_T = \sum_{j=1}^{p}\sum_{i=1}^{r}(x_{ij} - \bar{x})^2, \quad S_E = \sum_{j=1}^{p}\sum_{i=1}^{r}(x_{ij} - \bar{x}_j)^2,$$

$$S_A = r\sum_{j=1}^{p}(\bar{x}_j - \bar{x})^2, \quad S_T = S_E + S_A,$$

检验统计量

$$F = \frac{S_A/(p-1)}{S_E/(n-p)}$$

对显著性水平 α，当 $F > F_\alpha(p-1, n-p)$ 时，拒绝 H_0，即认为因

素 A 的影响显著;当 $F \leq F_\alpha(p-1, n-p)$ 时,接受 H_0,即认为因素 A 的影响不显著.

(七) 单因素方差分析中的参数估计

当拒绝 H_0 时,得到 μ_j 与 μ 的点估计:

$$\hat{\mu}_j = \overline{x}_j, \quad \hat{\mu} = \overline{x}.$$

两个总体 $N(\mu_j, \sigma^2)$ 和 $N(\mu_k, \sigma^2)(j \neq k)$ 均值差 $\mu_j - \mu_k$ 的区间估计:置信度为 $1-\alpha$ 的置信区间为

$$(\overline{x}_j - \overline{x}_k) \pm t_{\frac{\alpha}{2}}(n-p)\sqrt{\frac{2}{r} S_E/(n-p)}.$$

要 点 范 例

例 9.1 某建材实验室作陶粒混凝土实验,考察每立方米混凝土的水泥用量 x(单位:kg)对混凝土抗压强度 y(单位:kg/cm^2)的影响,测得数据如下表所示.

x	150	160	170	180	190	200	210	220	230	240	250	260
y	56.9	58.3	61.6	64.6	68.1	71.3	74.1	77.4	80.2	82.6	86.4	89.7

(1) 求经验回归方程 $\hat{y} = \hat{a} + \hat{b}x$;

(2) 检验一元线性回归的显著性($\alpha = 0.05$);

(3) 设 $x_0 = 225$,求 y 的预测值及置信度为 0.95 的预测区间.

解 (1) 由上表可得:$n = 12$, $\overline{x} = 205$, $\overline{y} = 72.6$, $l_{xx} = 14\ 300$, $l_{yy} = 1\ 323.82$, $l_{xy} = 4\ 347$.故

$$\hat{b} = \frac{l_{xy}}{l_{xx}} = \frac{4\ 347}{14\ 300} = 0.304;$$

$$\hat{a} = \overline{y} - \hat{b}\,\overline{x} = 72.6 - 0.304 \times 205 = 10.28.$$

得经验回归方程 $\hat{y} = 10.28 + 0.304x$.

(2) $U = \hat{b}l_{xy} = 0.304 \times 4\,347 = 1\,321.488$,

$$Q = l_{yy} - \hat{b}l_{xy} = 1\,323.82 - 1\,321.488 = 2.332,$$

$$F = (n-2)\frac{U}{Q} = 10 \times \frac{1\,321.488}{2.332} = 5\,666.76.$$

$\alpha = 0.05$, $F_{0.05}(1,10) = 4.96$,故回归方程显著.

$$\delta(225) = \sqrt{1 + \frac{1}{12} + \frac{(225-205)^2}{14\,300}} \approx 1.054,$$

(3) $s = \sqrt{\dfrac{Q}{n-2}} = \sqrt{\dfrac{2.332}{10}} = 0.482\,9$,

$$t_{0.025}(10) = 2.228\,1,$$

$$\hat{y}(225) = 10.28 + 0.304 \times 225 = 78.68.$$

而置信度为 0.95 的预测区间为:

$$78.68 \pm 2.228 \times 0.482\,9 \times 1.054 = 78.68 \pm 1.134,$$

即 $(77.546, 79.814)$.

例 9 2 某种合金钢的抗拉强度 y_1(单位:kg/ cm^2)和延伸率 y_2(单位:%)与钢中含碳量 x 有一定的关系.下表是这种合金钢 92 炉钢样记录数据合并后的数据表,要求分别建立指标 y_1, y_2 对于含碳量 x(单位:%)的回归方程,根据生产需要,此种合金钢的抗拉强度 y_1 应大于 32 千克/ 厘米^2,而延伸率 y_2 应大于 33%,这时若要以 95% 的把握满足上述要求,钢的含碳量应控制在什么范围内?

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x	0.05	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.16	0.18	0.20	0.21	0.23
y_1	40.8	41.7	41.9	42.8	42.0	43.6	44.8	45.6	45.1	48.9	50.0	55.0	54.8	60.0
y_2	39.9	38.5	37.8	39.1	39.6	36.1	40.1	38.5	40.6	35.3	37.0	32.3	34.2	32.4

解 (1) 求 y_1 对于 x 的回归方程 $\hat{y} = \hat{a}_1 + \hat{b}_1 x$.

$$\bar{x} = 0.133\,6, \quad \bar{y}_1 = 46.928\,6,$$

$$\sum_{i=1}^{14} x_i^2 = 0.2899, \quad \sum_{i=1}^{14} x_i y_{1i} = 91.909 .$$

$$l_{xy_1} = \sum_{i=1}^{14} x_i y_{1i} - 14 \times \bar{x} \times \bar{y} = 4.134,$$

$$l_{xx} = \sum_{i=1}^{14} x_i^2 - n\bar{x}^2 = 0.04,$$

$$\hat{b}_1 = \frac{l_{xy_1}}{l_{xx}} = 103.35, \quad \hat{a}_1 = \hat{y}_1 - \hat{b}_1 \bar{x} = 33.121 .$$

故 $\hat{y}_1 = 33.121 + 103.35x$.相关显著检验(F 检验法):

$$l_{yy_1} = \sum_{i=1}^{14} y_{1i}^2 - 14\bar{y}_1^2 = 31\,295.6 - 30\,832.1 = 463.53,$$

$$U_1 = \hat{b}_1 l_{xy_1} = 427.25,$$

$$Q_1 = l_{y_1y_1} - U_1 = 463.53 - 427.25 = 36.28,$$

$$F = \frac{U_1}{Q_1/(n-2)} = \frac{427.25}{36.28} \times 12$$

$$= 141.32 > F_{0.01}(1,12) = 9.33 .$$

故 y_1 对 x 线性相关显著 .

(2) 求 y_2 对 x 的回归直线方程 $\hat{y}_2 = \hat{a}_2 + \hat{b}_2 x$.

$$\bar{x} = 0.1336, \quad \bar{y}_2 = 37.2429,$$

$$\sum_{i=1}^{14} x_i y_{2i} = 68.067, \quad \sum_{i=1}^{14} x_i^2 = 0.2899,$$

$$l_{xy_2} = \sum_{i=1}^{14} x_i y_{2i} - 14 \times \bar{x} \times \bar{y} = -1.5731,$$

$$\hat{b}_2 = \frac{l_{xy_2}}{l_{xx}} = \frac{-1.5731}{0.04} = -39.3267,$$

$$\hat{a}_2 = \hat{y}_2 - \hat{b}_2 \bar{x} = 42.4969,$$

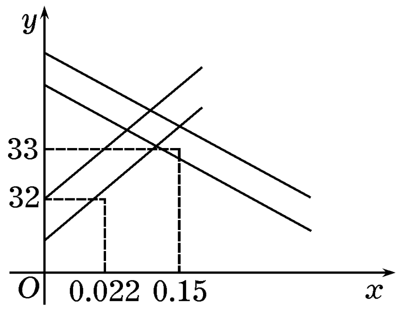
故 $\hat{y}_2 = 42.4969 - 39.3267x$.经检验 y_2 对 x 的线性相关也显著 .

(3) 求控制范围 .在置信度 95% 下求出预测直线 .

$$\hat{\sigma}_1 = \frac{Q_1}{n-2} = 1.739, \quad \hat{\sigma}_2 = \frac{Q_2}{n-2} = 1.808,$$

y_1: $\hat{y}_1 \pm \hat{\sigma}_1 u_{0.025} = 33.121 + 103.35x \pm 3.3376$,

y_2: $\hat{y}_2 \pm \hat{\sigma}_2 u_{0.025} = 42.4969 - 39.3267x \pm 3.6162$.



从左图可以看出,令 $33.121 + 103.35x - 3.3376 > 33$, $42.4969 - 39.3267x - 3.6162 > 33$ 就可满足条件,解出不等式

$$0.022\% < x < 0.15\%,$$

因此,只要把含碳量控制在 0.022% 和 0.15% 之间,则有 95% 的把握使合金钢抗拉强度 y_1 和延伸率 y_2 达到生产上的要求.

例 9.3 一批由同一种原料织成的布,用不同的印染工艺处理,然后进行缩水率(单位:%)试验,现采用5种不同的工艺,每种工艺处理4块布样,测得缩水率如下表所示.

工 艺	1	2	3	4
A_1	4.3	7.8	3.2	6.5
A_2	6.1	7.3	4.2	4.1
A_3	4.3	8.7	7.2	10.1
A_4	6.5	8.3	8.6	8.2
A_5	9.5	8.8	11.4	7.8

若布的缩水率服从正态分布,且经不同工艺处理的布的缩水率方差不同,试考察不同工艺对布的缩水率有无显著影响($\alpha = 0.05$).

解 经计算列表如下($p = 5$, $r = 4$).

工艺	$\sum_{i=1}^{r} x_{ij}$	$\sum_{i=1}^{r} x_{ij}^2$	$\left(\sum_{i=1}^{r} x_{ij}\right)^2$
A_1	21.8	131.82	475.24
A_2	21.7	124.95	470.89
A_3	30.3	248.03	918.09
A_4	31.6	252.34	998.56
A_5	37.5	358.49	1 406.25
总计	142.9	1 115.63	4 269.03

从而
$$P = \frac{(142.9)^2}{20} = 1\,021.020\,5.$$

$$Q = \frac{4\,269.03}{4} = 1\,067.257\,5, \quad R = 115.63,$$

$$S_E = R - Q = 1\,115.63 - 1\,067.257\,5 = 48.372\,5,$$

$$S_A = Q - P = 1\,067.257\,5 - 1\,021.020\,5 = 46.237\,0,$$

$$\overline{S}_A = \frac{S_A}{4} = 11.559\,3, \quad \overline{S}_E = \frac{S_E}{15} = 3.224\,8.$$

故
$$F = \frac{\overline{S}_A}{\overline{S}_E} = \frac{11.559\,3}{3.224\,8} = 3.584\,5,$$

$$\alpha = 0.05, \quad F_{0.05}(4,15) = 3.06.$$

因 $F > F_{0.05}(4,15)$,故认为不同工艺对缩水率有显著影响.

习 题 选 解

(9-1) 考察硫酸铜在水中的溶解度 y(单位:g)与温度 x(单位:℃)的关系时,做了 9 组实验,数据如下表所示.

x	0	10	20	30	40	50	60	70	80
y	14.0	17.5	21.2	26.1	29.2	33.3	40.0	48.0	54.8

求:(1) 经验回归方程 $\hat{y} = \hat{a} + \hat{b}x$;

(2) 相关系数 r;

(3) σ^2 在估计 $\hat{\sigma}^2$.

解 (1) 由数据可计算得 $\overline{x} = 40$, $\overline{y} = 31.567$,

$$l_{xx} = \sum_{i=1}^{n} x_i^2 - n(\overline{x})^2 = 20\,400 - 9 \times 40^2 = 6\,000,$$

$$l_{xy} = \sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y} = 14\,359 - 9 \times 40 \times 31.567 = 2\,995,$$

$$l_{yy} = \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2 = 10\,501.47 - 9 \times 31.567^2 = 1\,533.38.$$

故
$$\hat{b} = \frac{l_{xy}}{l_{xx}} = \frac{2\ 995}{6\ 000} = 0.499\ 2,$$
$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 31.567 - 0.499\ 2 \times 40 = 11.60,$$
得经验回归方程 $\hat{y} = 11.60 + 0.499\ 2x$.

$$(2)\quad r = \frac{l_{xy}}{\sqrt{l_{xx}}\sqrt{l_{yy}}} = \frac{2\ 995}{\sqrt{6\ 000}\sqrt{1\ 533.38}} \approx 0.987\ 4.$$

$$(3)\quad \hat{\sigma}^2 = \frac{Q}{n-2} = \frac{l_{yy} - \hat{b}l_{xy}}{n-2} \approx 5.468.$$

(9－2)　以家庭为单位,某种商品年需求量与该商品价格之间的一组调查数据如下表所示.

价格 x(元)	5	2	2	2.3	2.5	2.6	2.8	3	3.3	3.5
需求量 y(千克)	1	3.5	3	2.7	2.4	2.5	2	1.5	1.2	1.2

(1) **求经验回归方程** $\hat{y} = \hat{a} + \hat{b}x$;

(2) **检验线性关系的显著性**($\alpha = 0.05$).

解　(1) **由上表中数据可计算出**

$$\bar{x} = 2.9,\quad \bar{y} = 2.1,\quad l_{xx} = 7.18,\quad l_{yy} = 6.58,\quad l_{xy} = -5.93.$$

故
$$\hat{b} = \frac{l_{xy}}{l_{xx}} = \frac{-5.93}{7.18} = -0.826,$$
$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 2.1 + 0.826 \times 2.9 = 4.495.$$

经验回归方程为: $\hat{y} = 4.495 - 0.826x$.

$$(2)\quad U = \hat{b}l_{xy} = (-0.826) \times (-5.93) = 4.898,$$
$$Q = l_{yy} - \hat{b}l_{xy} = 6.58 - 4.898 = 1.682,$$
$$F = (n-1)\frac{U}{Q} = 8 \times \frac{4.898}{1.682} = 23.297,$$
$$\alpha = 0.05,\quad F_{0.05}(1,8) = 5.32,$$

因 $F > F_{0.05}(1,8)$,故线性关系是显著的.

(9－3)　通过原点的一元线性回归模型为:

$$y_i = bx_i + \varepsilon_i,\quad i = 1,2,\cdots,n.$$

其中 $\varepsilon_i (i = 1,2,\cdots,n)$ 相互独立,且都服从正态分布 $N(0,\sigma^2)$.

试由 n 组观察值$(x_i, y_i)(i = 1,2,\cdots, n)$,用最小二乘法估计 b.

解　记 $Q = \sum_{i=1}^{n} (y_i - bx_i)^2$,

$$\frac{\partial Q}{\partial b} = -2\sum_{i=1}^{n} (y_i - bx_i) x_i = 0,$$

从上面方程可解得

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

(9-4)　抽查某地区三所小学五年级男学生的身高,得数据如下:

小　学	身　高/cm					
第一小学	128.1	134.1	133.1	138.9	140.8	127.4
第二小学	150.3	147.9	136.8	126.0	150.7	155.8
第三小学	140.6	143.1	144.5	143.7	148.5	146.4

(1) 该地区三所小学五年级男学生的平均身高是否有显著差异($\alpha = 0.05$)?

(2) 分别求这三所小学五年级男学生的平均身高及方差的点估计.

解　(1) 计算结果如下表所示($p = 3$, $r = 6$).

小学	$\overline{x}_j$	$\sum_{i=1}^{r} (x_{ij} - \overline{x}_j)^2$
1	133.733 3	149.013 4
2	144.583 3	612.828 4
3	144.466 7	37.413 4

总计	422.7833	799.2552

$$S_E = 799.2552,\quad \bar{x} = 140.9278,$$

$$S_A = 6\sum_{i=1}^{3}(\bar{x}_j - \bar{x})^2 = 465.8840,$$

$$\bar{S}_E = \frac{S_E}{15} = \frac{799.2552}{15} = 53.2837,$$

$$\bar{S}_A = \frac{S_A}{2} = \frac{465.8840}{2} = 232.942,$$

$$F = \frac{\bar{S}_A}{\bar{S}_E} = \frac{232.942}{53.2837} = 4.3717.$$

查表得

$$F_{0.05}(2,15) = 3.68.$$

因 $F > F_{0.05}(2,15)$,故认为三所小学五年级男学生平均身高有显著差异.

(2) $\bar{\mu}_1 = 133.73,\quad \bar{\mu}_2 = 144.58,\quad \bar{\mu}_3 = 144.47,$

$$\bar{\sigma}^2 = \bar{S}_E = 53.28.$$

(9-5) 某化工厂在用钡泥制取硝酸钡的试验中,考虑到溶钡的溶出率随酸度的增大而提高.今将溶钡酸度从 pH = 4 降至 pH = 1,取 4 个水平:

$$A_1(\mathrm{pH}=4),\quad A_2(\mathrm{pH}=3),\quad A_3(\mathrm{pH}=2),\quad A_4(\mathrm{pH}=1).$$

每个水平做 4 次试验,测得硝酸钡含量如下表所示.

溶钡酸度	硝酸钡含量			
A_1	6.17	6.73	6.45	6.53
A_2	5.89	5.73	5.50	5.61
A_3	5.01	5.19	5.37	5.26
A_4	4.28	4.75	4.79	4.50

(1) 溶钡酸度对废水中的硝酸钡含量是否有显著影响 $(\alpha = 0.01)$?

(2) 求 μ_1、μ_2、μ_3、μ_4 及 σ^2 的点估计.

(3) 求 $\mu_1 - \mu_2$ 的置信度为 95% 的置信区间.

解 (1) 计算结果列表如下$(p = 4, r = 4)$.

溶钡酸度	$\overline{x}_j$	$\sum_{i=1}^{r}(x_{ij} - \overline{x}_j)^2$
A_1	6.47	0.161 6
A_2	5.682 5	0.083 9
A_3	5.207 5	0.068 5
A_4	4.58	0.169 4
总计	21.94	0.483 4

$$\overline{x} = \frac{21.94}{4} = 5.485, \quad S_E = 0.483\,4,$$

$$S_A = 4\sum_{j=1}^{4}(\overline{x}_j - \overline{x})^2 = 7.621\,0.$$

故

$$\overline{S}_A = 2.540\,3, \quad \overline{S}_E = 0.040\,3,$$

$$F = \frac{\overline{S}_A}{\overline{S}_E} = 63.034\,7, \quad F_{0.01}(3, 12) = 5.95.$$

因 $F > F_{0.01}(3, 12)$, 故认为溶钡酸度对废水中的硝酸钡含量有显著影响.

(2) $\hat{\mu}_1 = 6.47, \quad \hat{\mu}_2 = 5.86, \quad \hat{\mu}_3 = 5.21$

$\hat{\mu}_4 = 4.58, \quad \hat{\mu} = 5.49, \quad \hat{\sigma}^2 = 0.040.$

(3) $\mu_1 - \mu_2$ 的置信度为 95% 的置信区间为

$$(\overline{x}_1 - \overline{x}_2) \pm t_{\frac{\alpha}{2}}(n - p)\quad \frac{2}{r}\overline{S}_E$$

$$= (6.47 - 5.682\,5) \pm 2.178\,8 \times \quad \frac{2}{4} \times 0.040\,3$$

$$= 0.7875 \pm 0.3093.$$

近似为$(0.48, 1.10)$.

试 题 选 解

[9 - 1] 为研究温度 X 对某个化学过程的生产量 y 的影响,收集到如下数据:

X	-4	-3	-2	-1	0	1	2	3	4
Y	5	4	7	10	8	9	12	14	12

(1) 求 y 对 x 的线性回归方程 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

(2) 在误差 $\varepsilon \sim N(0, \sigma^2)$ 的假定下,检验回归方程的显著性($\alpha = 0.01$)(已知 $F_{0.99}(1,7) = 12.2$).

解 (1) 设 $y_i = \beta_0 + \beta_1 + x_i + \varepsilon_i$, $i = 1, 2, \cdots, 9$,

$$\varepsilon_i \sim N(0, \sigma^2), \text{且相互独立}.$$

由表中数据算得: $\overline{x} = 0$, $\overline{y} = 9$,

$$l_{xx} = \sum_{i=1}^{9} (x_i - \overline{x})^2 = 60,$$

$$l_{xy} = \sum_{i=1}^{9} (x_i - \overline{x})(y_i - \overline{y}) = 67,$$

$$l_{yy} = \sum_{i=1}^{9} (y_i - \overline{y})^2 = 90,$$

$$\hat{\beta}_1 = \frac{l_{xy}}{l_{xx}} = \frac{67}{60} = 1.1167,$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = \overline{y} = 9.$$

回归方程为 $$\hat{y} = 9 + 1.1167x.$$

(2) 要检验 $H_0: \beta_1 = 0$,

检验统计量 $$F = \frac{U}{Q_e / (n-2)}.$$

$$U = \frac{l_{xy}^2}{l_{xx}} = \frac{67^2}{60} \approx 74.82,$$

$$Q_e = l_{yy} - U = 90 - 74.82 = 15.18,$$

$$F = \frac{U}{Q_e/(n-2)} = \frac{74.82 \times 7}{15.18} = 34.5.$$

而 $$F_{1-0.01}(1,7) = F_{0.99}(1,7) = 12.2 < 34.5.$$

拒绝 H_0,即认为回归方程线性关系显著.

[9-2] 设有线性模型

$$y_1 = \beta_1 + \varepsilon_1;$$

$$y_2 = 2\beta_1 - \beta_2 + \varepsilon_2;$$

$$y_3 = \beta_1 + 2\beta_2 + \varepsilon_3;$$

$\varepsilon_1, \varepsilon_2, \varepsilon_3$ 独立同服从 $N(0, \sigma^2)$.

(1) 求 β_1, β_2 的最小二乘估计 $\hat{\beta}_1, \hat{\beta}_2$.

(2) 求 $\hat{\beta}_1, \hat{\beta}_2$ 的分布.

解 (1) 误差平方和

$$Q(\beta_1, \beta_2) = (y_1 - \beta_1)^2 + (y_2 - 2\beta_1 + \beta_2)^2 + (y_3 - \beta_1 - 2\beta_2)^2,$$

由

$$\frac{\partial Q}{\partial \beta_1} = -2(y_1 - \beta_1) - 4(y_2 - 2\beta_1 + \beta_2) - 2(y_3 - \beta_1 - 2\beta_2)$$

$$= 0;$$

$$\frac{\partial Q}{\partial \beta_2} = 2(y_2 - 2\beta_1 + \beta_2) - 4(y_3 - \beta_1 - 2\beta_2) = 0.$$

解得

$$\hat{\beta}_1 = \frac{1}{6}(y_1 + 2y_2 + y_3);$$

$$\hat{\beta}_2 = \frac{1}{5}(-y_2 + 2y_3).$$

(2) $$E(\hat{\beta}_1) = \frac{1}{6}E(y_1 + 2y_2 + y_3)$$

$$= \frac{1}{6}[\beta_1 + 2(2\beta_1 - \beta_2) + \beta_1 + 2\beta_2] = \beta_1,$$

即
$$E(\hat{\beta}_1)=\beta_1 .$$
$$D(\hat{\beta}_1)=\frac{1}{36}D(y_1+2y_2+y_3)=\frac{1}{36}(\sigma^2+4(\sigma^2)+\sigma^2)=\frac{\sigma^2}{6} .$$
同理可求出 $E(\hat{\beta}_2)=\beta_2$, $D(\hat{\beta}_2)=\dfrac{\sigma^2}{5}$,所以
$$\hat{\beta}_1 \sim N\left(\beta_1,\frac{\sigma^2}{6}\right),\quad \hat{\beta}_2 \sim N\left(\beta_2,\frac{\sigma^2}{5}\right) .$$

[9-3] 随机抽取某地区6个家庭的年收入 x 与年储蓄 y(单位:千元)资料如下表所示.

年收入 x	6	8	9	10	11	16
年储蓄 y	0.4	0.6	0.8	1.0	1.2	1.4

(1) 求 y 对 x 的线性回归方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x$.

(2) 在误差 $\varepsilon \sim N(0,\sigma^2)$ 的假定下,检验回归方程的显著性($\alpha=0.01$).

解　同[9-1]题类似方法可求得回归方程为
$$\hat{y}=-0.134+0.1034x$$
$$F=31>F_{0.99}(1,4)=21.2 .$$
拒绝 H_0,所得回归方程线性关系显著.

[9-4] 一位教师要检查三种不同数学方法的效果,为此随机地选取了水平相当的15位学生,把他们等分为三组.每组同一种教学方法,一段时间后,这位教师给这15位学生进行统考,成绩如下表所示.

方法	成绩				
甲	75	62	72	58	73
乙	81	84	68	92	90
丙	73	80	62	75	80

试问在显著性水平 $\alpha=0.05$ 下,这三种教学方法的效果有无显著性差异?这里假定第 i 组学生的成绩 $X_i \sim N(\mu_i,\sigma^2)$, $i=1,2,3$.(已知 $F_{0.95}(2,12)=3.89$)

解　要检验 $H_0: \mu_1 = \mu_2 = \mu_3$，

为了便于计算，将表中数据都减去 70，得 $y_{ij} = x_{ij} - 70$.

5	-8	2	-12	3
11	14	-2	22	20
3	10	-8	5	10

$$S_T = \sum_{i=1}^{r}\sum_{j=1}^{p} y_{ij}^2 - \frac{1}{n}\left(\sum_{i=1}^{r}\sum_{j=1}^{p} y_{ij}\right)^2$$

$$= 1\,749 - 375 = 1\,374,$$

$$S_A = \sum_{i=1}^{r}\frac{1}{p}\left(\sum_{j=1}^{p} y_{ij}\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{r}\sum_{j=1}^{p} y_{ij}\right)^2$$

$$= 945 - 375 = 570,$$

$$S_e = S_T - S_A = 804.$$

检验统计量

$$F = \frac{S_A/(r-1)}{S_e/(n-r)} = \frac{570/2}{804/12} = 4.254.$$

方差分析表为：

方差来源	平方和	自由度	均方	F 值
因素 A	570	2	285	4.254
误差	804	12	67	
总和	1 374	14		

因 $4.254 > 3.89 = F_{0.95}(2,12)$.

拒绝 H_0，认为三种教学方法的效果有显著性差异.

[9-5]　为了比较四种肥料对小麦亩产量的影响，取一块土壤肥沃程度和水利灌溉条件差不多的土地，分成 16 小块，将肥料品种记为 A_1，A_2，A_3，A_4，每种肥料施在四块土地上，得亩产量（单位：kg）如下表所示.

肥料品种	亩产量				$\overline{X}_i$	S_i^2
A_1	198	196	190	166	187.50	160.75
A_2	160	169	167	150	161.50	55.25

A_3	179	164	181	170	173.50	47.25
A_4	190	170	179	188	181.75	63.19

现算得总偏差平方和 $S_T = 2\ 540.94$.假定所给数据来自总体 $N(\mu_i, \sigma^2)$, $i = 1,2,3,4$.问在显著性水平 $\alpha = 0.05$ 下,施肥品种对小麦亩产量有无显著性影响($F_{0.95}(3,12) = 3.49$)?

解　设 X_i 为施肥料 A_i 所对应的亩产量,则

$$X_i \sim N(\mu_i, \sigma^2), \quad i = 1,2,3,4.$$

假设为　$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$.

经与[9-4]题类似方法计算得方差分析表如下所示.

方差分析表

方差来源	平方和	自由度	均方	F 值
肥料品种 A	1 527.10	3	509.06	4.65
误差	1 313.75	12	109.48	
总和	2 840.94	15		

$4.65 > 3.49 = F_{0.95}(3,12)$,拒绝 H_0,认为施肥种类对小麦亩产量有显著影响.

[9-6]　小白鼠在接种三种不同菌型伤寒杆菌后的存活日数如下表所示.

菌　形	存活日数				
Ⅰ	2	4	3	2	4
Ⅱ	5	6	8	6	10
Ⅲ	1	8	5	5	6

试问在显著性水平 $\alpha = 0.05$ 下,三种菌型的平均存活日数有无显著性差异?(假定存活日数服从正态分布且方差相等,已知 $F_{0.95}(2,12) = 3.89$).

解　这是单因素三水平重复试验，可用单因素的方差分析，由数据可计算出方差分析表如下所示．

方差分析表

方差来源	平方和	自由度	均方	F 值
因素	$S_A = 40$	2	20	5.2
误差	$S_e = 4.6$	12	3.83	
总和	$S_T = 86$	14		

因为 $5.2 > 3.89 = F_{0.95}(2, 12)$，所以认定三种菌型使小白鼠的平均存活日数有显著性差异．