第一章

习题 A

1．设 A、B、C 为三事件，用 A、B、C 及其运算关系表示下列事件.

(1) A 发生而 B 与 C 不发生;

(2) A、B、C 中恰好发生一个;

(3) A、B、C 中至少有一个发生;

(4) A、B、C 中恰好有两个发生;

(5) A、B、C 中至少有两个发生;

(6) A、B、C 中有不多于一个事件发生.

解：(1) $A\overline{B}\overline{C}$ 或 $A-B-C$ 或 $A-(B\cup C)$;

(2) $A\overline{B}\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\overline{B}C$;

(3) $A\cup B\cup C$ 或 $A\overline{B}\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\overline{B}C\cup AB\overline{C}\cup A\overline{B}C\cup\overline{A}BC\cup ABC$;

(4) $AB\overline{C}\bigcup A\overline{B}C\bigcup\overline{A}BC$.

(5) $AB\bigcup AC\bigcup BC$ 或 $AB\overline{C}\bigcup A\overline{B}C\bigcup\overline{A}BC\bigcup ABC$;

(6) $\overline{A}\overline{B}\overline{C}\bigcup A\overline{B}\overline{C}\bigcup\overline{A}B\overline{C}\bigcup\overline{A}\overline{B}C$.

2．一个样本空间有三个样本点， 其对应的概率分别为 $2p, p^2, 4p-1$， 求 p 的值.

解：由于样本空间所有的样本点构成一个必然事件，所以

$$2p+p^2+4p-1=1.$$

解之得 $p_1=-3+\sqrt{11}, p_2=-3-\sqrt{11}$，又因为一个事件的概率总是大于 0，所以 $p=-3+\sqrt{11}$.

3．已知 $P(A)=0.3$，$P(B)=0.5$，$P(A\cup B)=0.8$，求 (1) $P(AB)$; (2) $P(A-B)$;

(3) $P(\bar{A}\bar{B})$.

解：(1) 由 $P(A\cup B)=P(A)+P(B)-P(AB)$ 得

$$P(AB)=P(A)+P(B)-P(A\cup B)=03+0.5-0.8=0.$$

(2) $P(A-B)=P(A)-P(AB)=0.3-0=0.3$.

(3) $P(\bar{A}\bar{B})=1-P(\overline{\bar{A}\bar{B}})=1-P(A\cup B)=1-0.8=0.2.$

4. 设 $P(AB)=P(\bar{A}\bar{B})$，且 $P(A)=p$，求 $P(B)$.

解：由 $P(AB)=P(\bar{A}\bar{B})=1-P(\overline{\bar{A}\bar{B}})=1-P(A\cup B)=1-P(A)-P(B)+P(AB)$ 得

$P(A)+P(B)=1$，从而 $P(B)=1-p.$

5. 设 3 个事件 A、B、C，$P(A)=0.4$，$P(B)=0.5$，$P(C)=0.6$，$P(AC)=0.2$，

$P(BC)=0.4$ 且 $AB=\Phi$，求 $P(A\cup B\cup C)$.

解：

$$\begin{aligned}
&P(A\cup B\cup C)\\
&=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)\\
&=0.4+0.5+0.6-0-0.2-0.4+0\\
&=0.9.
\end{aligned}$$

6. 将 3 个球随机地放入 4 个杯子中去，求杯子中球的最大个数分别为 1，2，3 的概率.

解：依题意可知，基本事件总数为 4^3 个.

以 $A_i, i=1,2,3$ 表示事件“杯子中球的最大个数为 i”，则 A_1 表示每个杯子最多放一个球，共有 A_4^3 种方法，故

$$P(A_1)=\frac{A_4^3}{4^3}=\frac{6}{16}.$$

A_2 表示 3 个球中任取 2 个放入 4 个杯子中的任一个中，其余一个放入其余 3 个杯子中，放法总数为 $C_3^2C_4^1C_3^1$ 种，故

$$P(A_2)=\frac{C_3^2C_4^1C_3^1}{4^3}=\frac{9}{16}.$$

A_3表示 3 个球放入同一个杯子中，共有C_4^1种放法，故

$$P(A_3)=\frac{C_4^1}{4^3}=\frac{1}{16}.$$

7. 在整数 0 至 9 中任取 4 个，能排成一个四位偶数的概率是多少？

解：从 0 至 9 中任取 4 个数进行排列共有 $10\times9\times8\times7$ 种排法. 其中有$(4\times9\times8\times7-4\times8\times7+9\times8\times7)$种能成 4 位偶数.

故所求概率 $P=\dfrac{4\times9\times8\times7-4\times8\times7+9\times8\times7}{10\times9\times8\times7}=\dfrac{41}{90}$.

8. 一部五卷的文集，按任意次序放到书架上去，试求下列事件的概率：(1) 第一卷出现在旁边；(2) 第一卷及第五卷出现在旁边；(3) 第一卷或第五卷出现在旁边；(4) 第一卷及第五卷都不出现在旁边；(5) 第三卷正好在正中.

解：(1) 第一卷出现在旁边，可能出现在左边或右边，剩下四卷可在剩下四个位置上任意排，所以 $p=2\times4!/5!=2/5$.

(2) 可能有第一卷出现在左边而第五卷出现右边，或者第一卷出现在右边而第五卷出现在左边，剩下三卷可在中间三人上位置上任意排，所以 $p=2\times3!/5!=1/10$.

(3) $p=P${第一卷出现在旁边}+P{第五卷出现旁边}−P{第一卷及第五卷出现在旁边}$=\dfrac{2}{5}+\dfrac{2}{5}-\dfrac{1}{10}=\dfrac{7}{10}$.

(4) 这里事件是 (3) 中事件的对立事件，所以 $P=1-7/10=3/10$.

(5) 第三卷居中，其余四卷在剩下四个位置上可任意排，所以 $P=1\times4!/5!=1/5$.

9. 把 2，3，4，5 诸数各写在一张小纸片上，任取其三而排成自左向右的次序，求所得数是偶数的概率.

解：末位数可能是 2 或 4. 当末位数是 2(或 4)时，前两位数字从剩下四个数字中选排，所以 $P=2\times A_4^2/A_5^3=2/5$.

10. 一幢 10 层楼的楼房中的一架电梯，在底层登上 7 位乘客. 电梯在每一层

都停，乘客从第二层起离开电梯，假设每位乘客在哪一层离开电梯是等可能的，求没有两位及两位以上乘客在同一层离开的概率.

解：每位乘客可在除底层外的 9 层中任意一层离开电梯，现有 7 位乘客，所以样本点总数为 9^7. 事件 A“没有两位及两位以上乘客在同一层离开”相当于“从 9 层中任取 7 层，各有一位乘客离开电梯”. 所以包含 A_9^7 个样本点，于是 $P(A)=\frac{A_9^7}{9^7}$.

11. 两艘轮船都要停靠同一个泊位，它们可能在一昼夜的任意时刻到达. 设两船停靠泊位的时间分别为 1 小时与 2 小时，求有一艘船停靠泊位时必须等待一段时间的概率.

解：分别用 x,y 表示第一、二艘船到达泊位的时间. 一艘船到达泊位时必须等待当且仅当 $0\le x-y\le 2, 0\le y-x\le 1$. 因此所求概率为

$$P(A)=\frac{24^2-\frac{1}{2}\times 23^2-\frac{1}{2}\times 22^2}{24^2}\approx 0.121.$$

12. 10 个考签中有 4 个难签，3 个人参加抽签考试，不重复地抽取，每人一次，甲先，乙次，丙最后. 证明 3 人抽到难签的概率相同.

证明：设甲、乙、丙分别抽到难签的事件为 A,B,C，则，显然 $P(A)=\frac{4}{10}$.

$$P(B)=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})=\frac{4}{10}\times\frac{3}{9}+\frac{6}{10}\times\frac{4}{9}=\frac{4}{10}.$$

$$\begin{aligned}&P(C)\\&=P(AB)P(C|AB)+P(\bar{A}B)P(C|\bar{A}B)+P(A\bar{B})P(C|A\bar{B})+P(\bar{A}\bar{B})P(C|\bar{A}\bar{B})\\&=\frac{4}{10}\times\frac{3}{9}\times\frac{2}{8}+\frac{6}{10}\times\frac{4}{9}\times\frac{3}{8}+\frac{4}{10}\times\frac{6}{9}\times\frac{3}{8}+\frac{6}{10}\times\frac{5}{9}\times\frac{4}{8}\\&=\frac{4}{10}.\end{aligned}$$

13. 已知事件 A 发生的概率 $P(A)=0.5$，B 发生的概率 $P(B)=0.6$，以及条件概率 $P(B|A)$=0.8，求 A,B 和事件的概率.

解：由乘法公式得

$$P(AB)=P(A)P(B\mid A)=0.5\times 0.8=0.4.$$

所以

$$\begin{aligned}&P(A\cup B)\\&=P(A)+P(B)-P(AB)\\&=0.5+0.6-0.4=0.7.\end{aligned}$$

14. 一批零件共 100 个，其中次品有 10 个. 每次从中任取 1 个零件，取 3 次，取出后不放回. 求第 3 次才取得合格品的概率.

解：设 A_i 表示事件“第 i 次取得合格品”，则

$$P(\overline{A}_1\overline{A}_2A_3)=P(\overline{A}_1)P(\overline{A}_2\mid\overline{A}_1)P(A_3\mid\overline{A}_1\overline{A}_2)=\frac{10}{100}\times\frac{9}{99}\times\frac{90}{98}=\frac{9}{1078}\approx 0.00835.$$

15. 有两个袋子，每个袋子都装有 a 只黑球，b 只白球，从第一个袋中任取一球放入第二个袋中，然后从第二个袋中取出一球，求取得黑球的概率是多少？

解：设从第一个袋子摸出黑球 A，从第二个袋中摸出黑球为 B，则

$$P(A)=\frac{a}{a+b},\quad P(\overline{A})=\frac{b}{a+b},\quad P(B|A)=\frac{a+1}{a+b+1},\quad P(B|\overline{A})=\frac{a}{a+b+1},$$

由全概公式知：

$$P(B)=P(B|A)P(A)+P(B|\overline{A})P(\overline{A})=\frac{a}{a+b}.$$

16. 一个机床有 $\frac{1}{3}$ 的时间加工零件 A，其余时间加工零件 B. 加工零件 A 时，停机的概率是 0.3，加工零件 B 时，停机的概率时 0.4，求这个机床停机的概率.

解：设 C 表示“机床停机”，A 表示“加工零件 A”，B 表示“加工零件 B”，则

$$P(C)=P(A)P(C\mid A)+P(B)P(C\mid B)=\frac{1}{3}\times 0.3+\frac{2}{3}\times 0.4=\frac{11}{30}=0.367.$$

17. 现有两种报警系统 A 和 B，每种系统单独使用时，系统 A 有效的概率 0.92，系统 B 的有效概率为 0.93，在 A 失灵的条件下，B 有效的概率为 0.85，求

(1) 这两个系统至少有一个有效的概率；

(2) 在 B 失灵条件下，A 有效的概率.

解：设 A 表示“系统 A 有效”，B 表示“系统 B 有效”，则

$$P(A)=0.92, P(B)=0.93, P(B|\bar{A})=0.85.$$

由 $P(B|\bar{A})=\dfrac{P(B)-P(AB)}{1-P(A)}=0.85$ 知 $P(AB)=0.862$.

(1) $P(A\cup B)=P(A)+P(B)-P(AB)=0.92+0.93-0.862=0.988.$

(2) $P(A|\bar{B})=\dfrac{P(A)-P(AB)}{1-P(B)}=\dfrac{0.92-0.862}{1-0.93}=0.8285.$

18. 两部机器制造大量的同一种机器零件，根据长期资料总结，甲、乙机器制造出的零件废品率分别是 0.01 和 0.02. 现有同一机器制造的一批零件，估计这一批零件是乙机器制造的可能性比它们是甲机器制造的可能性大一倍，现从这批零件中任意抽取一件，经检查是废品. 试由此结果计算这批零件是由甲生产的概率.

解：设 A 表示“零件由甲生产”，B 表示“零件是次品”，则

$$P(A)=\frac{1}{3}, P(\bar{A})=\frac{2}{3}, P(B|A)=0.01, P(B|\bar{A})=0.02.$$

由贝叶斯公式有

$$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\bar{A})P(B|\bar{A})}=\frac{\frac{1}{3}\times 0.01}{\frac{1}{3}\times 0.01+\frac{2}{3}\times 0.02}=0.2.$$

19. 有朋友自远方来访，他乘火车、轮船、汽车、飞机来的概率分别是 0.3、0.2、0.1、0.4. 如果他乘火车、轮船、汽车来的话，迟到的概率分别是 $\frac{1}{4}$、$\frac{1}{3}$、$\frac{1}{12}$，而乘飞机则不会迟到. 结果他迟到了，试问他是乘火车来的概率是多少？

解：用 A_1 表示“朋友乘火车来”，A_2 表示“朋友乘轮船来”，A_3 表示“朋友乘汽车来”，A_4 表示“朋友乘飞机来”，B 表示“朋友迟到了”. 则

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{\sum_{k=1}^{4}P(A_k)P(B|A_k)}=\frac{1}{2}$$

20. 加工一个产品要经过三道工序，第一、二、三道工序不出现废品的概率分别是 0.9、0.95、0.8. 若假定各工序是否出废品相互独立，求经过三道工序而不出现废品的概率.

解：设 $A_i, i=1,2,3$ 分别表示第一、二、三道工序不出现废品，则

由独立性得

$P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3)=0.9\times0.95\times0.8=0.684.$

21. 三个人独立地破译一个密码，他们能译出的概率分别是 0.2、1/3、0.25. 求密码被破译的概率.

解：设 $A_i, i=1,2,3$ 分别表示第一、二、三个人破译出密码，则

由独立性得

$$\begin{aligned}&P(A_1\cup A_2\cup A_3)\\&=1-P(\overline{A_1\cup A_2\cup A_3})\\&=1-P(\overline{A_1}\,\overline{A_2}\,\overline{A_3})\\&=1-P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})\\&=1-0.8\times\frac{2}{3}\times0.75\\&=0.6.\end{aligned}$$

22. 对同一目标，3 名射手独立射击的命中率是 0.4、0.5 和 0.7，求三人同时向目标各射一发子弹而没有一发中靶的概率？

解：设 $A_i, i=1,2,3$ 分别表示第一、二、三个射手击中目标，则

由独立性得

$$P(\overline{A_1}\,\overline{A_2}\,\overline{A_3})=P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})=(1-0.4)(1-0.5)(1-0.7)=0.09.$$

23. 甲、乙、丙三人同时对飞机进行射击，三人击中的概率分别为 0.4、0.5、0.7. 飞机被一人击中而击落的概率为 0.2，被两人击中而击落的概率为 0.6，若三人都击中，飞机必定被击落，求飞机被击落的概率.

解：设 $C_i, i=1,2,3$ 依次表示甲、乙、丙击中飞机，$A_i, i=1,2,3$ 分别表示有 i 人击中飞机，B 表示飞机被击落，则

$$\begin{aligned}P(A_1)&=P(C_1\overline{C_2}\,\overline{C_3})+P(\overline{C_1}C_2\overline{C_3})+P(\overline{C_1}\,\overline{C_2}C_3)\\&=0.4\times0.5\times0.3+0.6\times0.5\times0.3+0.6\times0.5\times0.7\\&=0.06+0.09+0.21=0.36.\end{aligned}$$

$$P(A_2)=P(C_1C_2\overline{C_3})+P(C_1\overline{C_2}C_3)+P(C_1C_2\overline{C_3})$$
$$=0.4\times0.5\times0.3+0.4\times0.5\times0.7+0.6\times0.5\times0.7$$
$$=0.06+0.14+0.21=0.41.$$

$$P(A_3)=P(C_1C_2C_3)$$
$$=0.4\times0.5\times0.7$$
$$=0.14.$$

由全概率公式，得

$$P(B)=P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)$$
$$=0.36\times0.2+0.41\times0.6+0.14\times1$$
$$=0.458.$$

24. 观察四个新生儿的性别，设每一个出生婴儿是男婴还是女婴概率相等，求恰有 2 男 2 女的概率.

解：设 X 表示四个新生儿中男孩的个数，则由伯努利概型得

$$P(X=2)=C_4^2(\frac{1}{2})^2(\frac{1}{2})^2=\frac{3}{8}=0.375.$$

25. 灯泡使用寿命在 1000 小时以上的概率是 0.2，求 3 个灯泡在使用了 1000 小时后，最多坏了 1 个的概率.

解：设 A 表示灯泡使用在 1000 小时后坏了，则 $P(A)=p=0.8, q=1-p=0.2$.

设 X 表示三个灯泡中坏了的个数，则由伯努利概型得

$$P(X\le1)=P(X=0)+P(X=1)$$
$$=C_3^0p^0q^3+C_3^1p^1q^2$$
$$=0.2^3+3\times0.8\times0.2^2$$
$$=0.104.$$

26. 某机构有一个 9 人组成的顾问小组，若每个顾问贡献正确意见的概率是 0.7，现在该机构对某事可行与否个别征求各位顾问的意见，并按照多数人意见作出决策，求作出正确决策的概率.

解：设 X 表示贡献正确意见的人数，则由伯努利概型得

$$P(X\ge5)=\sum_{i=5}^{9}C_9^i0.7^i0.3^{9-i}$$
$$=C_9^5 0.7^5 0.3^4+C_9^6 0.7^6 0.3^3+C_9^7 0.7^7 0.3^2+C_9^8 0.7^8 0.3^1+C_9^9 0.7^9 0.3^0$$
$$=0.901.$$

习题 B

1．证明 $P(AB)+P(AC)-P(BC) \le P(A)$．

证：$P(AB)+P(AC)\text{-}P(BC)=p(AB\cup AC)+P(ABC)-P(BC)$

$$\le P(A)+P(BC)-P(BC)=P(A).$$

2．将 n 个人排成一排，甲与乙是其中两人，求这 n 个人的任意排列中，甲与乙之间恰有 r 个人的概率．如果 n 个人围成一圆圈，甲与乙之间恰有 r 个人的概率又是多少(在圆圈排列时仅考虑从甲到乙的顺时针方向)？

解：设 A，B 分别为排成一行或一圈

(1) $n=n!, m_A=C_2^1P_{n-2}^r(n-r-2+1)!\ P(A)=\dfrac{m_A}{n}=\dfrac{2(n-r-1)}{n(n-1)}$.

(2) $n=(n-1)!, m_B=C_{n-2}^r r!(n-r-2)!$，得 $P(B)=\dfrac{m_B}{n}=\dfrac{1}{n-1}$.

3. 15 个乒乓球中有 9 个新球，6 个旧球，第一次比赛取出了 3 个，用完了放回去，第二次比赛又取出 3 个，求第二次取出的 3 个球全是新球的概率.

解：设 A_i=第一次取出 i 个新球，$i=0,1,2,3$，B 表示第二次取出 3 个新球，则

$$P(B)=\sum_{i=0}^{3}P(A_i)P(B|A_i)=\frac{C_6^3}{C_{15}^3}\cdot\frac{C_9^3}{C_{15}^3}+\frac{C_6^2C_9^1}{C_{15}^3}\cdot\frac{C_8^3}{C_{15}^3}+\frac{C_6^1C_9^2}{C_{15}^3}\cdot\frac{C_7^3}{C_{15}^3}+\frac{C_9^3}{C_{15}^3}\cdot\frac{C_6^3}{C_{15}^3}=0.089.$$

4．证明：若三个事件 A、B、C 独立，则 $A\cup B$、AB 及 $A-B$ 都与 C 独立.

证明：(1) $P((A\cup B)C)=P(AC)+P(BC)-P(ABC)$

$$=P(A\cup B)P(C).$$

(2) $PABC)=P(A)P(B)P(C)=P(AB)P(C)$.

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

5．将线段 $(0,a)$ 任意折成三段，试求此三段线段构成三角形的概率.

解：设三条线段长度分别为 $x,y,a-x-y$，则样本空间为

$$\Omega=\{(x,y)\,|\,0<x<a,0<y<a,0<a-x-y<a\},$$

待求事件 A 满足

$$\begin{cases} x+y>a-x-y \\ x-y<a-x-y, \\ y-x<a-x-y \end{cases}$$

所以，$P(A)=\dfrac{L(A)}{L(\Omega)}=\dfrac{1}{4}$.

6. 要验收一批 100 件的物品，从中随机地取出 3 件来测试，设 3 件物品的测试是相互独立的，如果 3 件中有一件不合格，就拒绝接收该批物品. 设一件不合格的物品经测试查出的概率为 0.95，而一件合格品经测试误认为不合格的概率为 0.01，如果这 100 件物品中有 4 件是不合格的，问这批物品被接收的概率是多少？

解：设 A_i=抽到的 3 件物品中有 i 件不合格品，$i=0,1,2,3$. B=物品被接收，则

$$P(B)=\sum_{i=0}^{3}P(A_i)P(B\,|\,A_i)$$
$$=\frac{C_{96}^{3}}{C_{100}^{3}}\cdot 0.99^3+\frac{C_{96}^{2}C_4^1}{C_{100}^{3}}\cdot 0.99^2 0.05^1+\frac{C_{96}^{1}C_4^2}{C_{100}^{3}}\cdot 0.99^1 0.05^2+\frac{C_{96}^{0}C_4^3}{C_{100}^{3}}\cdot 0.99^0 0.05^3$$
$$=0.8629.$$

7. 一个人的血型为 O,A,B,AB 型的概率分别为 0.46、0.40、0.11、0.03，现在任意挑选五个人，求下列事件的概率

(1) 两个人为 O 型，其它三个人分别为其它三种血型；

(2) 三个人为 O 型，两个人为 A 型；

(3) 没有一人为 AB.

解 (1) 从 5 个人任选 2 人为 O 型，共有 $\binom{5}{2}$ 种可能，在其余 3 人中任选一人为 A 型，共有三种可能，在余下的 2 人中任选一人为 B 型，共有 2 种可能，另一人为 AB 型，顺此所求概率为：$\binom{5}{2}\times 3\times 2\times 0.46^2\times 0.40\times 0.11\times 0.13\approx 0.0168$

(2) $\binom{5}{3}\times 0.46^2\times 0.40^2\approx 0.1557$

(3) $(1-0.03)^5\approx 0.8587$.

8．设图 1-9 中两个系统中各元件通达与否相互独立，且每元件通达的概率均为 p，求系统 KL 与 KR 通达的概率．

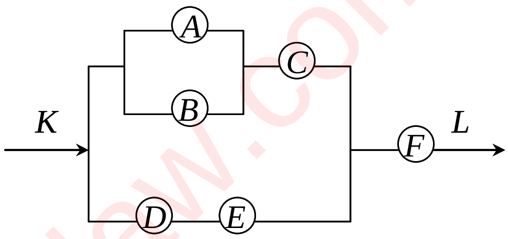


图 1-9 (a)

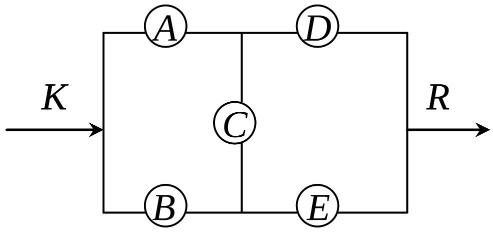


图 1-9 (b)

解： 设 $A^{'}, B^{'}$ 分别表示系统 KL 与 KR 通达，

(1) 解法一

$$
\begin{aligned}
&P(A^{'}) = P\{\{[A\cup B)\cap C]\cup(D\cap E)\}\cap F\} \\
&= P(ACF\cup BCF\cup DEF) \\
&= P(ACF)+P(BCF)+P(DEF)-P(ABCF)-P(ACDEF) \\
&-P(BCDEF)+P(ABCDEF) \\
&= p^3+p^3+p^3-p^4-p^5-p^5+p^6 \\
&= p^3(3-p-2p^2+p^3).
\end{aligned}
$$

解法二：

$$
\begin{aligned}
&P(A^{'}) = P\{\{[A\cup B)\cap C]\cup(D\cap E)\}\cap F\} \\
&= P(F)\{P[(A\cup B)\cap C]+P(DE)-P[(A\cup B)C(D\cap E)] \\
&= p[P(A\cup B)P(C)+P(D)P(E)-P(A\cup B)P(C)P(D)P(E)] \\
&= p[P(A)+P(B)-P(A)P(B)]p+p^3-p^4[P(A)+P(B)-P(AB)] \\
&= p^3(3-p-2p^2+p^3).
\end{aligned}
$$

(2)

$$
\begin{aligned}
&P(B^{'}) = P[\overline{C}(AD\cup BE)+(A\cup B)C(D\cup E)] \\
&=(1-p)(p^2+p^2-p^4)+(p+p-p^2)p(p+p-p^2) \\
&= p^2(2+2p-5p^2+2p^3).
\end{aligned}
$$

9．在伯努利试验中事件 A 出现的概率为 p，求在 n 重伯努利试验中事件 A 出现偶数次的概率．

解：设第 n 次出现事件 A 为 A，X 为 n 次实验中出现事件 A 的次数，则

$$
\begin{aligned}
&P_n(X=2k)\\
&=P_{n-1}(X=2k)P(\overline{A})+P_{n-1}(X=2k-1)P(A)\\
&=P_{n-1}(X=2k)(1-p)+p(1-P_{n-1}(X=2k))\\
&=P_{n-1}(X=2k)(1-2p)+p.
\end{aligned}
$$

令 $P_n(X=2k)=y_n$，则由如下差分方程

$$
\begin{cases} y_n-(1-2p)y_{n-1}=p \\ y_2=1-2p+2p^2 \end{cases}
$$

解之得 $P_n(X=2k)=y_n=\dfrac{1}{2}[1+(1-2p)^n].$

第 2 章

习题 A

1. 一个盒子中装有 10 个外形相同的球，其中白球 8 个，黑球 2 个. 如果从中取出一个球，若取到白球，则输 1 元，若取到黑球，则赢 4 元，记 X 表示赢钱数，求 X 的分布律.

解 由题意知，X 的可能取值为−1 和 4，且

$$P\{X=-1\}=\frac{8}{10}=0.8, \qquad P\{X=4\}=\frac{2}{10}=0.2$$

即

X	-1	4
p_k	0.8	0.2

2. 掷两个均匀的骰子，以 Y 表示这两次点数之和，试求随机变量 Y 的分布律.

解 样本空间为 $S=\{(1,1),(1,2),\cdots(1,6),\cdots,(6,1),(6,2),\cdots,(6,6)\}$，$S$ 中的样本数为 36.

随机变量 Y 的所有可能取值为 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

$\{Y=2\}=\{(1,1)\}$，$\{Y=3\}=\{(1,2),(2,1)\}$，$\{Y=4\}=\{(1,3),(2,2),(3,1)\}\cdots$

$\{Y=11\}=\{(5,6),(6,5)\}$，$\{Y=12\}=\{(6,6)\}$

于是 Y 的分布律为

Y	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

3. 从装有编号 1, 2, 3, 4, 5 五个球的袋中随机取出 3 个，记 X 为 3 个球中编号最大的一个，求随机变量 X 的分布律.

解 X 所有可能的取值为 3, 4, 5.

$P\{X=3\}=\dfrac{C_2^2}{C_5^3}=\dfrac{1}{10}$，(最大号码为 3，其余两球号为 1，2).

$P\{X=4\}=\dfrac{C_3^2}{C_5^3}=\dfrac{3}{10}$，(最大号码为 4，其余两球号为 1, 2, 3 中的两个).

$P\{X=3\}=\dfrac{C_4^2}{C_5^3}=\dfrac{6}{10}=\dfrac{3}{5}$，(最大号码为 3，其余两球号为 1, 2, 3, 4 中的两个).

于是 X 的分布律为

X	3	4	5
p_k	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{3}{5}$

4. 设某批电子元件的正品率为$\frac{4}{5}$，现对这批元件进行测试，只要测得一个正品元件，就停止测试工作，试求测试次数 X 的分布律.

解 测试次数 X 的所有可能取值为 1, 2, 3, ⋯. 当 $X=k$ 时，相当于前 $k-1$ 次测得的元件是次品，第 k 次是正品. 故 X 的分布律为

$$P\{X=k\}=(\frac{1}{5})^{k-1}\frac{4}{5}, k=1,2,\cdots.$$

5. 一个盒子中装有 5 个外形相同的球，其中 2 个白球，3 个黑球. 如果从中任取 3 个，X 表示其中白球的个数，求 X 的分布律.

解 X 的可能取值为 0, 1, 2. 且

$$P\{X=2\}=\frac{C_2^2C_3^1}{C_5^3}=\frac{3}{10}, P\{X=1\}=\frac{C_2^1C_3^2}{C_5^3}=\frac{6}{10}=\frac{3}{5}, P\{X=0\}=\frac{C_3^3}{C_5^3}=\frac{1}{10}.$$

于是 X 的分布律为

X	0	1	2
p_k	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{3}{10}$

6. 设随机变量 X 的分布律为 $P\{X=k\}=b\lambda^k, k=1,2,\cdots$，且 $b>0$ 为常数，求 λ.

解 由分布律的性质：$\sum_{k=1}^{+\infty}P\{X=k\}=1$ 可得

$$\sum_{k=1}^{+\infty}b\lambda^k=b\sum_{k=1}^{+\infty}\lambda^k=b\frac{\lambda}{1-\lambda}=1$$

从而 $\lambda=\frac{1}{1+b}$.

7. 设随机变量 ξ 服从参数为 λ 的泊松分布，且 $P\{\xi=1\}=P\{\xi=2\}$. 求 λ 和 $P\{\xi=3\}$.

解 因为随机变量 ξ 服从参数为 $\lambda(\lambda>0)$ 的泊松分布，所以

$$P\{\xi=k\}=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,\cdots$$

由 $P\{\xi=1\}=P\{\xi=2\}$，即

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

得 $\lambda=2$．从而

$$P\{\xi=3\}=\frac{2^3}{3!}e^{-2}=\frac{4}{3}e^{-2}$$

8. 已知某公司生产的螺丝以 0.001 的概率为次品，并设各个螺丝是否为次品是相互独立的．这家公司将每 100 个螺丝包成一包出售，并保证若发现某包内多于一个次品，则可退款．问卖出的各包螺丝中，被退回公司的占多大比例？(用泊松定理计算)

解　设 X 表示每包中次品的个数，则

$$P\{X\le 1\}=P\{X=0\}+P\{X=1\}$$

$$=C_{100}^{0}(0.001)^0(0.999)^{100}+C_{100}^{1}(0.001)^1(0.999)^{99}$$

$$\approx e^{-0.1}+0.1\cdot e^{-0.1}\approx 0.99532$$

从而被退回公司的比例为 0.00468．

9. 某商店出售某种商品，据历史记录分析，月销售量服从泊松分布，参数为 5，问在月初进货时要库存多少此种商品，才能以 0.999 的概率满足顾客的需要？

解　设 X 表示该种商品的月销售量，则由题意知 X 服从参数为 5 的泊松分布，其概率密度为

$$P\{\xi=k\}=\frac{5^k}{k!}e^{-5},k=0,1,2,\cdots$$

为保证以 0.999 的概率满足顾客的要求，该种商品月初的库存数 m 应当使得

$$P\{X\le m\}=0.999$$

即

$$P\{X>m\}=0.001$$

于是有

$$P\{X>m\}=\sum_{k=m+1}^{+\infty}P\{X=k\}=0.001$$

查泊松分布表，可得

$$m+1=14,m=13$$

故月初的库存至少有 13 件．

10. 设离散型随机变量 X 的分布律为 $P\{X=k\}=\dfrac{k}{10},k=1,2,3,4.$ 求

(1) $P\{1 \le X \le 2\}$；　(2) $P\left\{\frac{1}{2} \le X < 2\right\}$；　(3) 随机变量 X 的分布函数 $F(x)$.

解 (1) $P\{1 \le X \le 2\} = P\{X=1\} + P\{X=2\} = \frac{1}{10} + \frac{2}{10} = \frac{3}{10}$

(2) $P\left\{\frac{1}{2} \le X < 2\right\} = P\{X=1\} = \frac{1}{10}$

(3) X 仅在 1, 2, 3, 4 四点处的概率不为零, 而 $F(x)$ 的值是 $X \le x$ 的累积概率值, 由概率的有限可加性, 有

$$F(x)=\begin{cases} 0 & x<1 \\ P\{X=1\} & 1\le x<2 \\ P\{X=1\}+P\{X=2\} & 2\le x<3 \\ P\{X=1\}+P\{X=2\}+P\{X=3\} & 3\le x<4 \\ 1 & x\ge 4 \end{cases}$$

即

$$F(x)=\begin{cases} 0 & x<1 \\ \frac{1}{10} & 1\le x<2 \\ \frac{3}{10} & 2\le x<3 \\ \frac{3}{5} & 3\le x<4 \\ 1 & x\ge 4 \end{cases}$$

11. 设随机变量 X 的分布函数为

$$F(x)=\begin{cases} 0 & x<0 \\ \frac{1}{4} & 0\le x<1 \\ \frac{1}{2} & 1\le x<2 \\ 1 & x\ge 2 \end{cases}$$

求 X 的分布律.

解 由 $F(x)$ 的定义知

$$P\{X=a\} = F(a) - F(a-0)$$

于是

$P\{X=0\} = F(0) - F(0-0) = \frac{1}{4} - 0 = \frac{1}{4}$

$P\{X=1\}=F(1)-F(1-0)=\frac{1}{2}-\frac{1}{4}=\frac{1}{4}$

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

故 X 的分布律为

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

12. 如果 X 的一切可能取值为

$$(1)\left[0,\frac{\pi}{2}\right];\qquad (2)\left[-\frac{\pi}{2},\frac{\pi}{2}\right];\qquad (3)[0,\pi]$$

验证是 $f(x)=\sin x$ 否为连续型随机变量 ξ 的概率密度函数.

证明 (1) 因为在 $\left[0,\frac{\pi}{2}\right]$ 上, $\sin x\geq 0$, 并且 $\int_0^{\frac{\pi}{2}}\sin xdx=1$, 所以随机变量 X 可以由如下的概率密度

$$f(x)=\begin{cases}\sin x, & 0\leq x\leq\frac{\pi}{2}\\ 0, & \text{其它}\end{cases}$$

(2) 因为在区间 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ 上, $f(x)=\sin x$ 不是非负函数, 所以 $f(x)=\sin x$ 不是随机变量 X 的概率密度函数.

(3) 因为在区间 $[0,\pi]$ 上, $\int_0^{\pi}\sin xdx=2\neq 1$, 所以 $f(x)=\sin x$ 不是随机变量 X 的概率密度函数.

13. 设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<0\\ A\sin x, & 0\leq x<\frac{\pi}{2}\\ 1, & x\geq\frac{\pi}{2}\end{cases}$$

试求: (1) 常数 A; (2) $P\left\{-\frac{\pi}{3}<X<\pi\right\}$; (3) X 的概率密度函数 $f(x)$.

解 (1) 因为分布函数右连续,所以 $\lim_{x\to\frac{\pi}{2}+} F(x) = F(\frac{\pi}{2})$,得

$$A\sin\frac{\pi}{2} = 1 \Rightarrow A = 1$$

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

(3) 由分布函数和密度函数的关系,有

$$f(x) = F'(x) = \begin{cases} \cos x & 0 \le x \le \frac{\pi}{2} \\ 0 & \text{其它} \end{cases}$$

14. 已知随机变量 X 的分布函数为

$$F(x) = A + \frac{Be^x}{1+e^x}, -\infty < x < +\infty$$

试求:(1) 常数 A 和 B; (2) $P\{-1 < X < 1\}$; (3) X 的概率密度 $f(x)$.

解 (1) 由 $F(-\infty) = 0, F(+\infty) = 1$,可得

$$A = 0, B = 1$$

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

(3) $f(x) = F'(x) = \frac{e^x}{(1+e^x)^2}, -\infty < x < +\infty$

15. 设连续型随机变量 X 的概率密度为

$$f(x) = \begin{cases} A(1-x), & 0 < x < 1 \\ 0, & \text{其它} \end{cases}$$

试求:(1) 常数 A; (2) X 的分布函数 $F(x)$; (3) $P\left\{0 < X < \frac{1}{2}\right\}$; (4) $P\left\{\frac{1}{2} \le X < 2\right\}$.

解 (1) 由概率密度的性质 $\int_{-\infty}^{+\infty} f(x)dx = 1$,即有 $\int_0^1 A(1-x)dx = A - \frac{A}{2} = 1$,

可得 $A = 2$

(2) 由分布函数的定义

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

当 $x \le 0$ 时,显然 $F(x) = \int_{-\infty}^{x} f(t)dt = 0$;

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

当 $x\geq 1$ 时， $F(x)=\int_{-\infty}^{x} f(t)dt=\int_{0}^{1} 2(1-t)dt=1$

所以， $F(x)$ 的表达式为

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

(3) $P\left\{0<X<\frac{1}{2}\right\}=\int_{0}^{\frac{1}{2}} f(x)dx=\int_{0}^{\frac{1}{2}} 2(1-x)dx=(2x-x^2)\Big|_{0}^{\frac{1}{2}}=\frac{3}{4}$

(4) $P\left\{\frac{1}{2}\leq X<2\right\}=\int_{\frac{1}{2}}^{1} f(x)dx=\int_{\frac{1}{2}}^{1} 2(1-x)dx=(2x-x^2)\Big|_{\frac{1}{2}}^{1}=\frac{1}{4}$

16. 设连续型随机变量 X 的概率密度为

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

试求：(1) 常数 C; (2) X 的分布函数 $F(x)$; (3) $P\left\{-\frac{1}{2}<X<\frac{1}{2}\right\}$.

解 (1) 由概率密度函数性质 $\int_{-\infty}^{+\infty} f(x)dx=1$, 得

$$\int_{-\infty}^{+\infty} f(x)dx=\int_{-1}^{1}\frac{C}{\sqrt{1-x^2}}dx=C\cdot\arcsin x\,|_{-1}^{1}=C\pi=1$$

所以 $C=\dfrac{1}{\pi}$.

(2) 由分布函数的定义

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

当 $x\leq -1$ 时, 显然 $F(x)=\int_{-\infty}^{x} f(t)dt=0$;

当 $-1<x<1$ 时，$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-1}^{x}\frac{1}{\pi\sqrt{1-t^2}}dt=\frac{1}{\pi}\arcsin t\,|_{-1}^{x}=\frac{1}{\pi}\arcsin x+\frac{1}{2}$;

当 $x\geq 1$ 时， $F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-1}^{1}\frac{1}{\pi\sqrt{1-t^2}}dt=\frac{1}{\pi}\arcsin t\,|_{-1}^{1}=1$

所以， $F(x)$ 的表达式为

$$F(x)=\begin{cases}0 & x\le -1\\ \dfrac{1}{\pi}\arcsin x+\dfrac{1}{2} & -1<x<1\\ 1 & x\ge 1\end{cases}$$

17. 某公共汽车每隔 5 分钟有一辆汽车通过，乘客到达该汽车站的任一时刻是等可能的，求乘客候车时间不超过 3 分钟的概率.

解 设乘客到达时间为 X，则其概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{5}, & 0<x<5\\ 0, & 其它\end{cases}$$

乘客候车时间不超过 3 分钟，即到达时间在$[2,5]$之间. 因此，所求概率为

$$P\{2\le X\le 5\}=\int_2^5\frac{1}{5}dx=\frac{3}{5}$$

18. 设随机变量ξ在$[0,5]$上服从均匀分布，求方程$4x^2+4\xi x+\xi+2=0$有实根的概率.

解 由题意可知ξ的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{5}, & 0<x<5\\ 0, & 其它\end{cases}$$

而方程$4x^2+4\xi x+\xi+2=0$有实根的条件是

$$(4\xi)^2-4\cdot 4\cdot(\xi+2)\ge 0$$

即$\xi^2-\xi-2\ge 0$，

解得$\xi\ge 2$或$\xi\le -1$ （不在区间$[0,5]$上，故舍去）

因此，所求概率为

$$P\{\xi\ge 2\}=\int_2^{+\infty}f(x)dx=\int_2^5\frac{1}{5}dx=\frac{3}{5}$$

19. 修理一台机器所需时间（小时）服从参数为 2 的指数分布.

(1) 求修理时间超过 2 小时的概率；

(2) 已知修理时间已持续 9 小时的情况下，求修理时间至少是 10 小时的条件概率.

解 (1) 设修理时间为 X，则由题意知 X 的概率密度函数为

$$f(x)=\begin{cases}2e^{-2x} & x>0\\ 0 & x\le 0\end{cases}$$

因此，所求概率为

$$P\{X \ge 2\} = \int_2^{+\infty} f(x)dx = \int_2^{+\infty} 2e^{-2x}dx = (-e^{-2x})\Big|_2^{+\infty} = e^{-4}$$

(2) 由于指数分布具有无记忆性，故

$$P\{X \ge 10 \mid X \ge 9\} = P\{X \ge 1\}$$

$$= \int_1^{+\infty} f(x)dx = \int_1^{+\infty} 2e^{-2x}dx = (-e^{-2x})\Big|_1^{+\infty} = e^{-2}$$

20. 已知随机变量 X 的概率密度函数为 $f(x) = \frac{1}{2}e^{-|x|}, -\infty < x < +\infty$，求随机变量 X 的分布函数 $F(x)$.

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

当 $x \ge 0$ 时，$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} f(t)dt + \int_{0}^{x} f(t)dt = \frac{1}{2} + \int_{0}^{x} \frac{1}{2}e^{-t}dt = 1 - \frac{1}{2}e^{-x}$

所以

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

21. 设 ξ 服从正态分布 $N(0,1)$，求下列概率:

(1) $P\{\xi < 2.2\}$；(2) $P\{\xi > 1.6\}$；(3) $P\{\xi < -0.78\}$；(4) $P\{|\xi| < 1.55\}$；

(5) $P\{|\xi| > 2.5\}$.

解 (1) $P\{\xi < 2.2\} = \Phi(2.2) = 0.9861$

(2) $P\{\xi > 1.6\} = 1 - P\{\xi \le 1.6\} = 1 - \Phi(1.6) = 1 - 0.9608 = 0.0392$

(3) $P\{\xi < -0.78\} = \Phi(-0.78) = 1 - \Phi(0.78) = 0.2177$

(4) $P\{|\xi| < 1.55\} = P\{-1.55 < \xi < 1.55\} = \Phi(1.55) - \Phi(-1.55) = 0.8788$

(5) $P\{|\xi| > 2.5\} = P\{\xi > 2.5\} + P\{\xi < -2.5\} = 1 - \Phi(2.5) + \Phi(-2.5) = 0.0124$

22. 设 $X \sim N(\mu, \sigma^2)$，对

$$P\{\mu - k\sigma < X < \mu + k\sigma\} = 0.99$$

$$P\{\mu - k\sigma < X < \mu + k\sigma\} = 0.95$$

$$P\{\mu - k\sigma < X < \mu + k\sigma\} = 0.90$$

分别求出相应的 k 值。对于 k 取什么值有 $P\{X>\mu-k\sigma\}=0.95$.

解 因为 $X\sim N(\mu,\sigma^2)$，则有 $\frac{X-\mu}{\sigma}\sim N(0,1)$，故

$$P\{\mu-k\sigma<X<\mu+k\sigma\}=2\Phi(k)-1$$

故反查标准正态分布表得

当 $2\Phi(k)-1=0.99$ 时，即 $\Phi(k)=0.995$，得 $k=2.58$；

当 $2\Phi(k)-1=0.95$ 时，即 $\Phi(k)=0.975$，得 $k=1.96$；

当 $2\Phi(k)-1=0.90$ 时，即 $\Phi(k)=0.95$，得 $k=1.65$.

同理，由于

$$P\{X>\mu-k\sigma\}=1-P\{X\le\mu-k\sigma\}=1-\Phi(-k)=0.95$$

但 $\Phi(-k)=0.05<0.5$，无法反查表求 k，为此利用

$$\Phi(-k)=1-\Phi(k)$$

反查表得 $k=1.65$.

23. 某硬铝锻件的槽的宽度服从 $\mu=0.9000$ 与 $\sigma=0.0030$ 为参数的正态分布. 规定的限度是 0.9000 ± 0.0050. 试求：

(1) 这种锻件的废品率.

(2) 如果这个随机变量有以 $\mu=0.9000$ 与 σ 为参数的正态分布，为使 100 个产品中废品不多于 1 个，可允许的最大 σ 值.

解 (1) 设 X 表示锻件的槽的宽度，由题意可得 $X\sim N(0.9000,0.0030^2)$，

$$P\{0.8950<X<0.9050\}=2\Phi\left(\frac{5}{3}\right)-1\approx0.9050$$

故这种锻件的废品率为 0.0950.

(2) $P\{0.8950<X<0.9050\}=2\Phi\left(\frac{0.0050}{\sigma}\right)-1\ge0.99$

可得最大 σ 值为 0.0019.

24. 某地区的降雨量近似服从均值为 40.2cm，标准差为 8.4cm 的正态分布.

(1) 明年的降雨量超过 44cm 的概率是多少？

(2) 以后 7 年中恰有 3 年的降雨量超过 44cm 的概率是多少？

解 (1) 设 X 表示明年的降雨量，由题意知 $X\sim N(40.2,8.4^2)$，于是

$$P\{X>44\}=1-P\{X\le 44\}=1-\Phi\left(\frac{3.8}{8.4}\right)=0.3264$$

(2) 以后 7 年中恰有 3 年的降雨量超过 44cm 的概率为

$$C_7^3(0.3264)^3(0.6736)^4=0.2506$$

25. 设随机变量 X 在区间[2,5]上服从均匀分布,求对 X 进行三次独立观测中,至少有两次的观测值大于 3 的概率.

解 由题意,随机变量 X 的概率密度为

$$f(x)=\begin{cases}\frac{1}{3}, & 2\le x\le 5\\ 0, & \text{其它}\end{cases}$$

从而 $P\{X>3\}=\int_3^5 \frac{1}{3}dx=\frac{2}{3}$

令 Y 表示对 X 进行三次独立观测中,观测值大于 3 的次数,则所求概率为

$$P\{Y\ge 2\}=C_3^2(\frac{2}{3})^2\frac{1}{3}+C_3^3(\frac{2}{3})^3=\frac{4}{9}+\frac{8}{27}=\frac{20}{27}$$

26. 某仪器有三只独立工作的同型号电气元件，其寿命（以小时为单位）都服从同一指数分布，密度函数为

$$f(x)=\begin{cases}\frac{1}{600}e^{-\frac{x}{600}}, & x>0\\ 0, & x\le 0\end{cases}$$

试求在仪器使用的最初 200 个小时内,至少有一个电子元件损坏的概率.

解 用 $A_i(i=1,2,3)$ 表示事件“在仪器使用的最初 200 个小时内,第 i 个元件损坏”,

$X_i(i=1,2,3)$ 表示第 i 个元件的寿命. 则有

$$P(A_i)=P(X_i\le 200)=\int_0^{200}\frac{1}{600}e^{-\frac{x}{600}}dx=1-e^{-\frac{1}{3}}$$

于是 $P(\bar{A}_i)=e^{-\frac{1}{3}}$

因此,所求概率为

$$P(A_1\cup A_2\cup A_3)=1-P(\bar{A}_1\bar{A}_2\bar{A}_3)=1-(e^{-\frac{1}{3}})^3=1-e^{-1}$$

27. 测量一目标的距离时,发生的随机误差 ξ (米) 具有概率密度

$$f(x)=\frac{1}{40\sqrt{2\pi}}e^{-\frac{(x-20)^2}{3200}},-\infty<x<+\infty$$

求三次测量中,至少一次误差的绝对值不超过 30 米的概率.

解 已知ξ服从正态分布$X \sim N\left(20, 40^2\right)$，从而

$$P\{|\xi|<30\}=P\{-30<\xi<30\}$$
$$=\Phi\left(\frac{30-20}{40}\right)-\Phi\left(\frac{-30-20}{40}\right)$$
$$=\Phi(0.25)-\Phi(-1.25)$$
$$=0.5897-0.1056=0.4931$$

所以三次测量中，至少一次误差的绝对值不超过 30 米的概率为

$$p=1-(P\{|\xi|>30\})^3=1-(1-P\{|\xi|\le 30\})^3$$
$$=1-(1-0.4913)^3 \approx 0.87$$

28. 设随机变量 X 的概率分布为

X	-1	0	1	2	3
p_k	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

求：(1) $X-1$；(2) X^2的概率分布.

解 (1) $X-1$的分布律为

$X-1$	-2	-1	0	1	2
p_k	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

(2) X^2的分布律为

X	0	1	4	9
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$

29. 设随机变量 X 在$(-1,1)$上服从均匀分布，求$Y=X^2$的概率密度$f_Y(y)$.

解 由题意，X 的概率密度为

$$f(x)=\begin{cases}\frac{1}{2}, & -1<x<1 \\ 0, & \text{其它}\end{cases}$$

当$y<0$时，有$F_Y(y)=P\{Y\le y\}=P\{X^2\le y\}=0$；

当 $0 \le y < 1$ 时，有

$$F_Y(y) = P\{Y \le y\} = P\left\{X^2 \le y\right\} = P\left\{-\sqrt{y} \le X \le \sqrt{y}\right\}$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} dx = \sqrt{y}$$

当 $y \ge 1$ 时，有

$$F_Y(y) = P\{Y \le y\} = P\left\{X^2 \le y\right\} = P\left\{-\sqrt{y} \le X \le \sqrt{y}\right\}$$

$$= \int_{-\sqrt{y}}^{-1} 0 dx + \int_{-1}^{1} \frac{1}{2} dx + \int_{1}^{\sqrt{y}} 0 dx = 1$$

于是，Y 的分布函数为

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \sqrt{y} & 0 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$

求导得 Y 的概率密度为

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} & 0 \le y < 1 \\ 0 & \text{其它} \end{cases}$$

30. 设随机变量 X 在 $(-1,1)$ 上服从均匀分布，求 $Y = |X|$ 的概率密度 $f_Y(y)$.

解 由题意，X 的概率密度为

$$f(x) = \begin{cases} \dfrac{1}{2}, & -1 < x < 1 \\ 0, & \text{其它} \end{cases}$$

当 $y < 0$ 时，有 $F_Y(y) = P\{Y \le y\} = P\left\{X^2 \le y\right\} = 0$；

当 $0 \le y < 1$ 时，有

$$F_Y(y) = P\{Y \le y\} = P\left\{|X| \le y\right\} = P\{-y \le X \le y\}$$

$$= \int_{-y}^{y} \frac{1}{2} dx = y$$

当 $y \ge 1$ 时，有

$$F_Y(y) = P\{Y \le y\} = P\left\{|X| \le y\right\} = P\{-y \le X \le y\}$$

$$= \int_{-y}^{-1} 0 dx + \int_{-1}^{1} \frac{1}{2} dx + \int_{1}^{y} 0 dx = 1$$

于是，Y 的分布函数为

$$F_Y(y)=\begin{cases}0 & y<0\\ y & 0\le y<1\\ 1 & y\ge 1\end{cases}$$

求导得 Y 的概率密度为

$$f_Y(y)=\begin{cases}1 & 0\le y<1\\ 0 & 其它\end{cases}$$

31. 设随机变量 X 的概率密度为

$$f_X(x)=\begin{cases}e^{-x}, & x\ge 0\\ 0, & x<0\end{cases}$$

求随机变量 $Y=e^X$ 的概率密度 $f_Y(y)$.

解 因为 $y=e^x$ 单调增加, 其反函数为

$$x=h(y)=\ln y \qquad (y>0)$$

故

$$f_Y(y)=f_X(\ln y)\left|\ln' y\right|=f_X(\ln y)\cdot\frac{1}{y}$$

$$=\begin{cases}e^{-\ln y}\cdot\dfrac{1}{y}, & \ln y\ge 0\\ 0, & \ln y<0\end{cases}$$

$$=\begin{cases}\dfrac{1}{y^2}, & y\ge 1\\ 0, & y<1\end{cases}$$

32. 设 θ 服从 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 上的均匀分布, 求 $R=A\sin\theta$ 的概率密度, 其中 A 是固定的常数. 这样的随机变量 R 出现在弹道理论中, 若在与地面成 α 角处将发射物点火, 那么落点 R 可以表示成 $R=\frac{v^2}{g}\cdot\sin 2\alpha$, 其中 g 是重力常数, 等于 $980cm/s^2$.

解 由题意知 θ 的概率密度为

$$f_X(x)=\begin{cases}\dfrac{1}{\pi}, & -\dfrac{\pi}{2}<x<\dfrac{\pi}{2}\\ 0, & 其它\end{cases}$$

由于 $R=A\sin\theta$ 在 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 上单调, 且其反函数为

$$\theta = \arcsin\frac{R}{A} \qquad (-A < R < A)$$

从而 $R = A\sin\theta$ 的概率密度为

$$f_Y(y) = f_X(\arcsin\frac{y}{A})\frac{1}{\sqrt{A^2 - y^2}}$$

$$= \begin{cases} \dfrac{1}{\pi}\dfrac{1}{\sqrt{A^2-y^2}}, & -\dfrac{\pi}{2} < \arcsin\dfrac{y}{A} < \dfrac{\pi}{2} \\ 0, & \text{其它} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\pi\sqrt{A^2-y^2}}, & -A < y < A \\ 0, & \text{其它} \end{cases}$$

习题 B

1. 假定掷一个骰子 2 次. 试求下列随机变量的分布律:
(1) 两次掷出的最大点数;
(2) 两次掷出的最小点数;
(3) 两次掷出点数之积;
(4) 第一次掷出的点数减去第二次掷出的点数.

解 (1)

Y	1	2	3	4	5	6
p_k	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

(2)

Y	1	2	3	4	5	6
p_k	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

(3)

Y	1	2	3	4	5	6	8	9	10	12	15
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Y	16	18	20	24	25	30	36
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

(4)

Y	−5	−4	−3	−2	−1	0	1	2	3	4	5

p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2. 从装有 8 个白球,4 个黑球与 2 个黄球的箱中随机地取出 2 个球.假定每取出 1 个黑球赢 2 美元,而取出 1 个白球输 1 美元.以 X 表示赢的钱数,求 X 的分布律.

解 由题意知,X 的可能取值为 4,2,1,0,−1,−2.且

$$P\{X=4\}=\frac{C_4^2}{C_{14}^2}=\frac{6}{91}$$

$$P\{X=2\}=\frac{C_4^1C_2^1}{C_{14}^2}=\frac{8}{91}$$

$$P\{X=1\}=\frac{C_4^1C_8^1}{C_{14}^2}=\frac{32}{91}$$

$$P\{X=0\}=\frac{C_2^2}{C_{14}^2}=\frac{1}{91}$$

$$P\{X=-1\}=\frac{C_2^1C_4^1}{C_{14}^2}=\frac{8}{91}$$

$$P\{X=-2\}=\frac{C_8^2}{C_{14}^2}=\frac{28}{91}$$

于是 X 的分布律为

X	4	2	1	0	−1	−2
p_k	$\frac{6}{91}$	$\frac{8}{91}$	$\frac{32}{91}$	$\frac{1}{91}$	$\frac{8}{91}$	$\frac{28}{91}$

3. 设随机变量 X 与 Y 均服从正态分布 $X\sim N(\mu,4^2),Y\sim N(\mu,5^2)$，记 $p_1=P\{X\le\mu-4\}$， $p_2=P\{Y\ge\mu+5\}$ 则

A. 对任何实数 μ，都有 $p_1=p_2$.　　B. 对任何实数 μ，都有 $p_1<p_2$.

C. 对任何实数 μ，都有 $p_1>p_2$.　　D. 只对个别的个别值 μ，才有 $p_1=p_2$.

解 选 A.

$$p_1=P\{X\le\mu-4\}=\Phi(-1)$$

$$p_2=P\{Y\ge\mu+5\}=1-P\{Y<\mu+5\}=1-\Phi(1)$$

而$1-\Phi(1)=\Phi(-1)$

所以 $p_1=p_2$.

4. 设随机变量 $X\sim N(\mu,\sigma^2)$, 则随 σ 的增大, 概率 $P\{|X-\mu|\le\sigma\}$ （ ）

A. 单调增大　B. 单调减小　C. 保持不变　D. 增减不定

解 选 C.

$$P\{|X-\mu|\le\sigma\}=P\{\mu-\sigma\le X\le\mu+\sigma\}=2\Phi(1)-1$$

为常数.

5. 设随机变量 X 的概率密度 $f(x)$ 为偶函数, 而 $F(x)$ 是 X 的分布函数, 则对任意实数 a, 有（ ）

A. $F(-a)=1-\int_0^a f(x)dx$　B. $F(-a)=\frac{1}{2}-\int_0^a f(x)dx$

C. $F(-a)=F(a)$　D. $F(-a)=2F(a)-1$

解 选 B.

因为 $1=\int_{-\infty}^{+\infty}f(x)dx=2\int_0^{+\infty}f(x)dx$, 所以 $\int_0^{+\infty}f(x)dx=\frac{1}{2}$.

而

$F(-a)=\int_{-\infty}^{-a}f(x)dx=\int_{+\infty}^{a}f(-t)(-dt)=\int_a^{+\infty}f(x)dx$

$=\int_0^{+\infty}f(x)dx-\int_0^a f(x)dx=\frac{1}{2}-\int_0^a f(x)dx$

6. 设 X 和 Y 是任意两个相互独立的连续性随机变量, 他们的概率密度分别为 $f_1(x)$ 和 $f_2(x)$, 分布函数分别为 $F_1(x)$ 和 $F_2(x)$. 则

A. $f_1(x)+f_2(x)$ 必为某一随机变量的概率密度.

B. $F_1(x)F_2(x)$ 必为某一随机变量的分布函数.

C. $F_1(x)+F_2(x)$ 必为某一随机变量的分布函数.

D. $f_1(x)f_2(x)$ 必为某一随机变量的概率密度.

解 选 B.

由已知, $\int_{-\infty}^{+\infty}f_1(x)dx=\int_{-\infty}^{+\infty}f_2(x)dx=1$, 故

$$\int_{-\infty}^{+\infty}[f_1(x)+f_2(x)]dx=\int_{-\infty}^{+\infty}f_1(x)dx+\int_{-\infty}^{+\infty}f_2(x)dx=2\neq1$$

所以不选 A.

由分布函数的性质知 $F_1(+\infty)=1, F_2(+\infty)=1$, 故 $\lim\limits_{x\to+\infty}[F_1(x)+F_2(x)]=2\neq1$,

所以不选 C.
若设

$$f_1(x)=f_2(x)=\begin{cases}\dfrac{1}{2} & 0<x<2\\ 0 & \text{其它}\end{cases}$$

则

$$f_1(x)\cdot f_2(x)=\begin{cases}\dfrac{1}{4} & 0<x<2\\ 0 & \text{其它}\end{cases}$$

这时

$$\int_{-\infty}^{+\infty} f_1(x)\cdot f_2(x)dx=\int_0^2\frac{1}{4}dx=\frac{1}{2}\neq 1$$

所以不选 D.

若令 $g(x)=F_1(x)F_2(x)$, 由 $F_1(-\infty)=0, F_2(-\infty)=0, F_1(+\infty)=1, F_2(+\infty)=1$,

可得 $g(-\infty)=0, g(+\infty)=1$,

又由于 $F_1(x)$ 和 $F_2(x)$ 非降, 可得 $g(x)=F_1(x)F_2(x)$ 非降.

再由 $F_1(x)$ 和 $F_2(x)$ 右连续, 可见 $g(x)=F_1(x)F_2(x)$ 也是右连续.

所以 $g(x)=F_1(x)F_2(x)$ 是分布函数.

7. 设随机变量 X 服从参数为 $(2,p)$ 的二项分布, 随机变量 Y 服从参数为 $(3,p)$ 的二项分布, 若 $P\{X\geq 1\}=\frac{5}{9}$, 求 $P\{Y\geq 1\}$.

解 因为 X 服从参数为 $(2,p)$ 的二项分布, 且 $P\{X\geq 1\}=\frac{5}{9}$, 所以 $P\{X=0\}=\frac{4}{9}$.

即 $C_2^0p^0(1-p)^2=\frac{4}{9}$. 得 $p=\frac{1}{3}$.

从而 $P\{Y\geq 1\}=1-P\{Y=0\}=1-C_3^0p^0(1-p)^3=1-\frac{8}{27}=\frac{19}{27}$.

8. 设乘客 10: 00 到达公共汽车站, 已知公共汽车到站的时间服从 10: 00 到 10: 30 间的均匀分布. 试求:

(1) 必须等 10 分钟以上的概率;

(2) 若 10: 15 时, 公共汽车仍未到, 乘客至少还要再等 10 分钟的概率.

解 (1) 令 X 表示公共汽车 10: 00 后到达车站所需时间, 由于 X 在 (0, 30) 上服从均匀分布, 所以乘客等 10 分钟以上当且仅当汽车在 10: 10 分到 10: 30 之间到站. 因此, 所求概率为

$$P\{10<X<30\}=\int_{10}^{30}\frac{1}{30}dx=\frac{2}{3}$$

(2) $P\{X>25|X>15\}=\dfrac{P\{X>25,X>15\}}{P\{X>15\}}=\dfrac{P\{X>25\}}{P\{X>15\}}=\dfrac{\frac{1}{6}}{\frac{1}{2}}=\dfrac{1}{3}$

9. 设顾客在某银行的窗口等待时间 X 服从指数分布,其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{5}e^{-\frac{x}{5}}, & x>0\\ 0, & x\le 0\end{cases}$$

某顾客在窗口等待服务,若超过 10min,他就离开,他一个月要到银行 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 $P\{Y\ge 1\}$.

解 由题意知, $P\{X>10\}=\int_{10}^{+\infty}\frac{1}{5}e^{-\frac{1}{5}x}dx=e^{-2}$, $P\{X\le 10\}=1-e^{-2}$.

于是 Y 的分布律为

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

从而

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

10. 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{3}, & 0<x<1\\ \dfrac{2}{9}, & 3<x<6\\ 0, & 其它\end{cases}$$

若 k 使得 $P\{X\ge k\}=\dfrac{2}{3}$, 求 k 的取值范围.

解 若 $k\le 0$, 则

$$P\{X\ge k\}=\int_k^{+\infty}f(x)dx=\int_0^1\frac{1}{3}dx+\int_3^6\frac{2}{9}dx=1>\frac{2}{3}$$

故 $k>0$,

当 $0<k<1$ 时,

$$P\{X\ge k\}=\int_k^{+\infty}f(x)dx=\int_k^1\frac{1}{3}dx+\int_3^6\frac{2}{9}dx=\frac{1}{3}(1-k)+\frac{2}{3}=\frac{2}{3}$$

故 $k=1$.

当 $1\le k\le 3$ 时,

$$P\{X\ge k\}=\int_k^{+\infty}f(x)dx=\int_3^6\frac{2}{9}dx=\frac{2}{3}$$

故对任意 $1\le k\le 3$, 均成立.

当 $3<k\le 6$ 时,

$$P\{X \geq k\} = \int_k^{+\infty} f(x)dx = \int_k^6 \frac{2}{9}dx = \frac{2}{9}(6-k) < \frac{2}{3}$$

故对任意$3<k\leq 6$,均不成立.
当$k>6$时,

$$P\{X \geq k\} = 0$$

综上,k 的取值范围为$[1,3]$.

11. 设某电子管寿命ξ服从参数为$\mu=160$的正态分布,若$P(-120<\xi<200)=0.8$,可允许的σ最大为多少?

解 $P(-120<\xi<200) = \Phi\left(\frac{200-160}{\sigma}\right) - \Phi\left(\frac{120-160}{\sigma}\right)$

$$= \Phi\left(\frac{40}{\sigma}\right) - \Phi\left(\frac{-40}{\sigma}\right) = 2\Phi\left(\frac{40}{\sigma}\right) - 1 = 0.8$$

所以$\Phi\left(\frac{40}{\sigma}\right)=0.9$,查表得$\frac{40}{\sigma}\approx 1.28$,所以$\sigma\approx 31.25$.

12. 设随机变量 X 服从参数为 2 的指数分布,证明: $Y=1-e^{-2X}$在区间(0, 1)上服从均匀分布.
证明 由题意知,X 的概率密度为

$$f_X(x) = \begin{cases} 2e^{-2x}, & x>0 \\ 0, & x \leq 0 \end{cases}$$

又因为$y=1-e^{-2x}$单调递增,且反函数为

$$x = -\frac{1}{2}\ln(1-y) \qquad (0<y<1)$$

故

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

$$= \begin{cases} \frac{1}{1-y} e^{-2\cdot[-\frac{1}{2}\ln(1-y)]}, & -\frac{1}{2}\ln(1-y) > 0 \\ 0, & -\frac{1}{2}\ln(1-y) \leq 0 \end{cases}$$

$$= \begin{cases} 1, & 0<y<1 \\ 0, & \text{其它} \end{cases}$$

第 3 章

习题 A

1. 某包中有 5 个晶体管, 其中 2 个是次品. 每次从中取出一个进行检验, 直到所有的次品都检验出来为止. 记 X 为直到查出第一个次品时所需要的检验次数, Y 为查出第一个次品后再查出第二个次品所需要的检验次数. 求 X 和 Y 的联合概率密度.

解 (X, Y) 所有可能的取值为

(1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5). 且

$$P\{X=1,Y=2\}=\frac{2}{5}\cdot\frac{1}{4}=\frac{1}{10}$$

$$P\{X=1,Y=3\}=\frac{2}{5}\cdot\frac{3}{4}\cdot\frac{1}{3}=\frac{1}{10}$$

$$P\{X=1,Y=4\}=\frac{2}{5}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{10}$$

$$P\{X=1,Y=5\}=\frac{2}{5}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}\cdot 1=\frac{1}{10}$$

$$P\{X=2,Y=3\}=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{1}{3}=\frac{1}{10}$$

$$P\{X=2,Y=4\}=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{10}$$

$$P\{X=2,Y=5\}=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}\cdot 1=\frac{1}{10}$$

$$P\{X=3,Y=4\}=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{1}{10}$$

$$P\{X=3,Y=5\}=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}\cdot 1=\frac{1}{10}$$

$$P\{X=4,Y=5\}=\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{1}{3}\cdot 1\cdot 1=\frac{1}{10}$$

于是 X 和 Y 的联合分布律为

$Y \backslash X$	1	2	3	4
2	$\frac{1}{10}$	0	0	0
3	$\frac{1}{10}$	$\frac{1}{10}$	0	0
4	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	0
5	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

2. 设随机变量(X, Y)具有分布函数

$$F(x,y)=\begin{cases}(1-e^{-x})(1-e^{-y}) & x>0,y>0\\ 0 & \text{其它}\end{cases}$$

求边缘分布函数.

解 由 $\lim\limits_{y\to+\infty}F(x,y)=F_X(x)$, $\lim\limits_{x\to+\infty}F(x,y)=F_Y(y)$, 可得

$$F_X(x)=\begin{cases}1-e^{-x} & x>0\\ 0 & x\le 0\end{cases},\quad F_Y(y)=\begin{cases}1-e^{-y} & y>0\\ 0 & y\le 0\end{cases}$$

3. 将一枚硬币掷3次, 以X表示前2次中出现H的次数, 以Y表示3次中出现H的次数, 求X和Y的联合分布律以及(X, Y)的边缘分布律.

解 (X, Y)所有可能的取值为(0, 0), (0, 1), (1, 1), (1, 2), (2, 2), (2, 3). 且

$$P\{X=0,Y=0\}=(\frac{1}{2})^3=\frac{1}{8}$$

$$P\{X=0,Y=1\}=(\frac{1}{2})^2\cdot\frac{1}{2}=\frac{1}{8}$$

$$P\{X=1,Y=1\}=C_2^1(\frac{1}{2})(\frac{1}{2})\cdot\frac{1}{2}=\frac{1}{4}$$

$$P\{X=1,Y=2\}=C_2^1(\frac{1}{2})(\frac{1}{2})\cdot\frac{1}{2}=\frac{1}{4}$$

$$P\{X=2,Y=2\}=C_2^2(\frac{1}{2})^2\cdot\frac{1}{2}=\frac{1}{8}$$

$$P\{X=2,Y=3\}=C_2^2(\frac{1}{2})^2\cdot\frac{1}{2}=\frac{1}{8}$$

于是, X和Y的联合分布律为

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

4. 设X和Y的联合概率密度函数为

$$f(x,y)=\frac{6}{7}(x^2+\frac{xy}{2})\qquad 0<x<1,0<y<2$$

求: (1) $P\{X>Y\}$; (2) $P\left\{Y>\frac{1}{2}\mid X<\frac{1}{2}\right\}$

解 (1) $P\{X>Y\}=\iint\limits_{x>y} f(x,y)dxdy=\int_0^1 dx\int_0^x \frac{6}{7}(x^2+\frac{xy}{2})dy$

$$=\int_0^1 \frac{6}{7}(x^3+\frac{x^2}{4})dx=\frac{15}{56}$$

(2) $P\left\{Y>\frac{1}{2}\,|\,X<\frac{1}{2}\right\}=\dfrac{P\left\{Y>\frac{1}{2},X<\frac{1}{2}\right\}}{P\left\{X<\frac{1}{2}\right\}}$

$$=\frac{\int_0^{\frac{1}{2}}dx\int_{\frac{1}{2}}^2 \frac{6}{7}(x^2+\frac{xy}{2})dy}{\int_0^{\frac{1}{2}}dx\int_0^2 \frac{6}{7}(x^2+\frac{xy}{2})dy}=\frac{69}{80}$$

5. 设 X 和 Y 的联合概率密度为

$$f(x,y)=c\left(y^2-x^2\right)e^{-y} \qquad -y<x<y,0<y<+\infty$$

求：(1) 常数 c. (2) 边缘概率密度.

解 (1) 由性质 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy=1$，可得

$$\int_0^{+\infty}dy\int_{-y}^{y}c(y^2-x^2)e^{-y}dx=c\int_0^{+\infty}e^{-y}(2y^3-\frac{2}{3}y^3)dy$$

$$=\frac{4}{3}c\int_0^{+\infty}y^3e^{-y}dy=8c=1$$

所以 $c=\frac{1}{8}$.

(2) 当 $x<0$ 时，

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-x}^{+\infty}\frac{1}{8}\left(y^2-x^2\right)e^{-y}dy$$

$$=\frac{7x^2-16x+16}{64}e^{x}$$

当 $x\geq 0$ 时，

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{x}^{+\infty}\frac{1}{8}\left(y^2-x^2\right)e^{-y}dy$$

$$=\frac{7x^2+16x+16}{64}e^{-x}$$

从而

$$f_X(x)=\begin{cases}\dfrac{7x^2-16x+16}{64}e^{x} & x<0\\ \dfrac{7x^2+16x+16}{64}e^{-x} & x\ge 0\end{cases}$$

同理

当 $y\le 0$ 时，

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=0$$

当 $y>0$ 时，

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

$$=\int_{-y}^{y}\frac{1}{8}\left(y^2-x^2\right)e^{-y}dx=\frac{1}{6}y^3e^{-y}$$

从而

$$f_Y(y)=\begin{cases}0 & y\le 0\\ \dfrac{1}{6}y^3e^{-y} & y>0\end{cases}$$

6. 从数集 $\{1,2,3,4,5\}$ 中随机地选出一个数 X，再从不大于 X 的子集中即 $\{1,2,\cdots,X\}$ 中随机地选取一个数，记次数为 Y.

求：(1) 求 X 和 Y 的联合分布律；

(2) 对 $i=1,2,3,4,5$，在给定 $Y=i$ 的条件下 X 的条件分布律.

解 (1) X 和 Y 的联合分布律为

X \ Y	1	2	3	4	5
1	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{20}$	$\frac{1}{25}$
2	0	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{20}$	$\frac{1}{25}$
3	0	0	$\frac{1}{15}$	$\frac{1}{20}$	$\frac{1}{25}$
4	0	0	0	$\frac{1}{20}$	$\frac{1}{25}$
5	0	0	0	0	$\frac{1}{25}$

(2) 对任意的$1 \le k \le i$，有条件分布律

$$P\{Y = k \mid X = i\} = \frac{1}{i}, i = 1,2,3,4,5.$$

7. 设 X 和 Y 的联合概率密度为

$$f(x,y) = xe^{-x(y+1)} \qquad x > 0, y > 0$$

求在给定$Y = y(y > 0)$的条件下，X 的条件概率密度.

解 在给定$Y = y(y > 0)$的条件下，X 的条件概率密度为

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \dfrac{xe^{-x(y+1)}}{\int_0^{+\infty} xe^{-x(y+1)}dx} & x > 0 \\ 0 & \text{其它} \end{cases}$$

$$= \begin{cases} \dfrac{xe^{-x(y+1)}}{(y+1)^2}, & x > 0, \\ 0, & \text{其它}. \end{cases}$$

8. 设 X 和 Y 的联合密度函数为

$$f(x,y) = \frac{1}{8}(x^2 - y^2)e^{-x} \qquad x > 0, -x < y < x$$

求在给定$X = x$的条件下，Y 的条件概率密度.

解 当$x \le 0$时，$f_{Y|X}(y \mid x) = 0$.

当$x > 0$时，$f_{Y|X}(y \mid x) = \dfrac{f(x,y)}{f_X(x)} = \dfrac{\frac{1}{8}(x^2 - y^2)e^{-x}}{\int_{-x}^{+x} \frac{1}{8}(x^2 - y^2)e^{-x}dy} = \dfrac{3(x^2 - y^2)}{4x^3}$

故

$$f_{Y|X}(y \mid x) = \begin{cases} \dfrac{3(x^2 - y^2)}{4x^3} & x > 0 \\ 0 & x \le 0 \end{cases}$$

9. 设(X,Y)在中心为(0,0)，边长为 2 的正方形区域上服从均匀分布.

(1) 证明 X 与 Y 是独立的，且都服从(−1,1)上的均匀分布.

(2) 求$P\{X^2 + Y^2 \le 1\}$.

(1)证明 由题意知，X 和 Y 的联合概率密度为

$$f(x,y) = \begin{cases} \dfrac{1}{4} & -1 < x < 1, -1 < y < 1 \\ 0 & \text{其它} \end{cases}$$

从而当 $x \le -1$ 或 $x \ge 1$ 时， $f(x,y)=0$，有 $f_X(x)=0$.

当 $-1<x<1$ 时， $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\int_{-1}^{1}\frac{1}{4}dy=\frac{1}{2}$

因此
$$f_X(x)=\begin{cases}\frac{1}{2} & -1<x<1\\ 0 & \text{其它}\end{cases}$$

同理

$$f_Y(y)=\begin{cases}\frac{1}{2} & -1<y<1\\ 0 & \text{其它}\end{cases}$$

从而 $f_X(x)f_Y(y)=f(x,y)$，故 X 与 Y 是独立的，且都服从 $(-1,1)$ 上的均匀分布.

(2) $P\left\{X^2+Y^2\le 1\right\}=\iint\limits_{x^2+y^2\le 1} f(x,y)dxdy=\iint\limits_{x^2+y^2\le 1}\frac{1}{4}dxdy=\frac{\pi}{4}$

10. 设 X 和 Y 的联合密度函数为

$$f(x,y)=\begin{cases}x+y & 0<x<1,0<y<1\\ 0 & \text{其它}\end{cases}$$

(1) 问 X 和 Y 是否独立？ (2) 求 $P\{X+Y<1\}$.

解当 $x\le 0$ 或 $x\ge 1$ 时， $f(x,y)=0$，有 $f_X(x)=0$.

当 $0<x<1$ 时， $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\int_0^1 (x+y)dy=x+\frac{1}{2}$

因此
$$f_X(x)=\begin{cases}x+\frac{1}{2} & 0<x<1\\ 0 & \text{其它}\end{cases}$$

同理

$$f_Y(y)=\begin{cases}y+\frac{1}{2} & 0<y<1\\ 0 & \text{其它}\end{cases}$$

从而 $f_X(x)f_Y(y)\ne f(x,y)$，故 X 与 Y 不是相互独立的，

(2) $P\{X+Y<1\}=\int_0^1 dx\int_0^{1-x}(x+y)dy=\frac{1}{3}$

11. 设 A, B 为独立随机变量，在 (0, 1) 上都服从均匀分布. 求:

(1) A 和 B 的联合概率密度函数.

(2) 方程 $Ax^2+Bx+1=0$ 的根全是实根的概率.

解 (1) A 和 B 的联合概率密度函数为

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

(2) 方程 $Ax^2+Bx+1=0$ 的根全是实根等价于 $B^2\geq 4A$.

故所求概率为

$$p=\iint\limits_{y^2\geq 4x} f(x,y)dxdy=\int_0^1 dy\int_0^{\frac{y^2}{4}} dx=\frac{1}{12}$$

12. 一个男人和一个女人约定在下午 14:30 在某地见面. 如果男人到达的时间在 14:15 到 14:45 之间服从均匀分布, 女人到达的时间在 14:00 到 15:00 之间服从均匀分布, 且两个人到达的时间是相互独立的. 试求先到者等待时间不超过 5 分钟的概率.

解 令 X 表示男人到达的时间, Y 表示女人到达的时间, 于是 X 和 Y 的联合概率密度为

$$f(x,y)=\begin{cases}\dfrac{1}{1800} & 15<x<45,0<y<60\\ 0 & \text{其它}\end{cases}$$

所求概率为

$$P\{|X-Y|\leq 5\}=\int_{15}^{45}dx\int_{x-5}^{x+5}\frac{1}{1800}dy=\frac{1}{6}$$

13. 设 X 在 (0, 1) 上服从均匀分布, Y 服从参数为 1 的指数分布, 又设 X 和 Y 相互独立, 分别求: (1) $Z=X+Y$; (2) $Z=Y/X$ 的概率密度.

解 因为 X 在 (0, 1) 上服从均匀分布, 所以 X 的概率密度为

$$f_X(x)=\begin{cases}1 & 0<x<1\\ 0 & \text{其它}\end{cases}$$

因为 Y 服从参数为 1 的指数分布, 所以 Y 的概率密度为

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

由 X 和 Y 相互独立, 故 X 和 Y 的联合概率密度为

$$f(x,y)=\begin{cases}e^{-y} & 0<x<1,y>0\\ 0 & \text{其它}\end{cases}$$

(1) 当 $z<0$ 时, 由于 $f(x,y)=0$, 所以 $F(z)=P\{X+Y\leq z\}=0$.

当 $0\leq z\leq 1$ 时, $F(z)=P\{X+Y\leq z\}=\int_0^z dx\int_0^{z-x}e^{-y}dy=\int_0^z(1-e^{x-z})dx=z-1+e^{-z}$.

当 $z>1$ 时，$F(z)=P\{X+Y\le z\}=\int_0^1 dx\int_0^{z-x} e^{-y}dy=\int_0^1(1-e^{x-z})dx=1-(e-1)e^{-z}$.

故

$$F(z)=\begin{cases} 0 & z<0 \\ z-1+e^{-z} & 0\le z\le 1 \\ 1-e^{1-z}+e^{-z} & z>1 \end{cases}$$

求导，得

$$f(z)=F'(z)=\begin{cases} 0 & z<0 \\ 1-e^{-z} & 0\le z\le 1 \\ e^{1-z}-e^{-z} & z>1 \end{cases}$$

(2) 当 $z\le 0$ 时，由于 $f(x,y)=0$，所以 $F(z)=P\left\{\frac{Y}{X}\le z\right\}=0$.

当 $z>0$ 时，$F(z)=P\left\{\frac{Y}{X}\le z\right\}=\int_0^1 dx\int_0^{zx} e^{-y}dy=\int_0^1(1-e^{-xz})dx=1+\frac{1}{z}e^{-z}-\frac{1}{z}$.

故

$$F(z)=\begin{cases} 0 & z\le 0 \\ 1+\frac{1}{z}e^{-z}-\frac{1}{z} & z>0 \end{cases}$$

求导，得

$$f(z)=F'(z)=\begin{cases} 0 & z\le 0 \\ -\frac{1}{z^2}e^{-z}-\frac{1}{z}e^{-z}+\frac{1}{z^2} & z>0 \end{cases}$$

14. 设随机变量 X 和 Y 相互独立，且 $X\sim E(1),Y\sim E(2)$，即 X 和 Y 分别服从参数为 1 和 2 的指数分布. 求：$Z=X+2Y$ 的概率密度.

解 由题意知，X 和 Y 的联合概率密度为

$$f(x,y)=\begin{cases} 2e^{-(x+2y)} & x>0,y>0 \\ 0 & \text{其它} \end{cases}$$

解 $Z=X+2Y$ 的分布函数为

$$F(z)=P\{Z\le z\}=P\{X+2Y\le z\}=\iint_{x+2y\le z} f(x,y)dxdy$$

当 $z\le 0$ 时，由于 $f(x,y)=0$，所以 $F(z)=P\{X+2Y\le z\}=0$.

当 $z>0$ 时，

$$F(z)=P\{X+2Y\le z\}=\int_0^{\frac{z}{2}} 2e^{-2y}dy\int_0^{z-2y} e^{-x}dx.$$

$$=\int_0^{\frac{z}{2}} 2e^{-2y}-2e^{-z}dy=1-e^{-z}-ze^{-z}$$

故

$$F(z)=\begin{cases} 0 & z\le 0 \\ 1-e^{-z}-ze^{-z} & z>0 \end{cases}$$

求导, 得

$$f(z)=F'(z)=\begin{cases} 0 & z\le 0 \\ ze^{-z} & z>0 \end{cases}$$

习题 B

1. 设在某一天里进入邮局的人数是一个参数为 λ 的泊松随机变量. 证明: 如果一个进入邮局的人是男性的概率为 p, 是女性的概率为 $1-p$, 则进入邮局的男性人数和女性人数是分别以 λp 和为参数 $\lambda(1-p)$ 的独立泊松随机变量.

解 令 X 与 Y 分别表示进入邮局的男性人数和女性人数. 为得到 $P\{X=i,Y=j\}$ 的表达式, 将 $X+Y$ 作为条件如下:

$$\begin{aligned}P\{X=i,Y=j\}&=P\{X=i,Y=j\mid X+Y=i+j\}P\{X+Y=i+j\}\\&\quad+P\{X=i,Y=j\mid X+Y\ne i+j\}P\{X+Y\ne i+j\}\end{aligned}$$

由于 $P\{X=i,Y=j\mid X+Y\ne i+j\}$ 为 0, 所以得到

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

现在, 因为 $X+Y$ 代表进入邮局的总人数, 故由假设可知

$$P\{X+Y=i+j\}=e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}$$

此外, 考虑 $i+j$ 个人进入邮局, 因为每个进入邮局的人是男性的概率是 p, 所以恰有个人为男性的概率正好是二项概率 $C_{i+j}^{i}p^{i}(1-p)^{j}$. 也就是说

$$P\{X=i,Y=j\mid X+Y=i+j\}=C_{i+j}^{i}p^{i}(1-p)^{j}$$

故

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

$$=e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}C_{i+j}^{i}p^{i}(1-p)^{j}$$

$$=e^{-\lambda}\frac{(\lambda p)^i}{i!\,j!}[\lambda(1-p)]^j$$

$$=e^{-\lambda p}\frac{(\lambda p)^i}{i!}e^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!}$$

从而

$$P\{X=i\}=e^{-\lambda p}\frac{(\lambda p)^i}{i!}\sum_j e^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!}=e^{-\lambda p}\frac{(\lambda p)^i}{i!}$$

同理可得

$$P\{Y=j\}=e^{-\lambda(1-p)}\frac{[\lambda(1-p)]^j}{j!}$$

2. 设 X 和 Y 分别以为 λ_1,λ_2 参数的独立泊松随机变量，证明：$Z=X+Y$ 是以 $\lambda_1+\lambda_2$ 为参数的泊松分布.

解 $P\{X+Y=n\}=\sum_{k=0}^{n}P\{X=k,Y=n-k\}=\sum_{k=0}^{n}P\{X=k\}P\{Y=n-k\}$

$$=\sum_{k=0}^{n}e^{-\lambda_1}\frac{\lambda_1^{\ k}}{k!}e^{-\lambda_2}\frac{\lambda_2^{\ n-k}}{(n-k)!}=e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^{\ k}\cdot\lambda_2^{\ n-k}}{k!(n-k)!}$$

$$=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^{\ k}\cdot\lambda_2^{\ n-k}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n$$

即证 $Z=X+Y$ 是以 $\lambda_1+\lambda_2$ 为参数的泊松分布.

3. 设 X 和 Y 分别是参数为 (s,λ) 和 (t,λ) 的独立 Γ 随机变量，则 X 和 Y 的概率密度函数分布为

$$f_X(x)=\begin{cases}\dfrac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x>0\\ 0 & x\le 0\end{cases};\qquad f_Y(y)=\begin{cases}\dfrac{\lambda e^{-\lambda y}(\lambda y)^{t-1}}{\Gamma(t)} & y>0\\ 0 & y\le 0\end{cases}$$

证明：$Z=X+Y$ 也是 Γ 随机变量，其参数为 $(s+t,\lambda)$.

证明： 由卷积公式

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

当 $z\le 0$ 时，有

$$f_Z(z)=0.$$

当 $z>0$ 时,有

$$f_Z(z)=\int_0^z \frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)}\cdot\frac{\lambda e^{-\lambda(z-x)}[\lambda(z-x)]^{t-1}}{\Gamma(t)}dx$$

$$=Ke^{-\lambda z}\int_0^z x^{s-1}(z-x)^{t-1}dx$$

其中 K 为常数,令 $y=\frac{x}{z}$,

$$=Ce^{-\lambda z}z^{s+t-1}\int_0^1 y^{s-1}(1-y)^{t-1}dy$$

$$=Ce^{-\lambda z}z^{s+t-1}$$

其中 C 是不依赖于 z 的常数.因为是密度函数,所以积分为 1,可以确定 C,有

$$f_Z(z)=\frac{\lambda e^{-\lambda z}(\lambda z)^{s+t-1}}{\Gamma(s+t)}$$

即

$$f_Z(z)=\begin{cases}\dfrac{\lambda e^{-\lambda z}(\lambda z)^{s+t-1}}{\Gamma(s+t)} & z>0\\ 0 & z\le 0\end{cases}$$

第 4 章

习题 A

1. 设随机变量 X 的分布律为

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

求：$E(X)$，$E(X^2+2)$ 及 $D(X)$.

解 由期望的定义，可得 $E(X)=0\cdot\frac{1}{4}+1\cdot\frac{1}{2}+2\cdot\frac{1}{4}=1$，

$$E(X^2)=0^2\cdot\frac{1}{4}+1^2\cdot\frac{1}{2}+2^2\cdot\frac{1}{4}=\frac{3}{2}$$

从而 $E(X^2+2)=E(X^2)+2=\frac{3}{2}+2=\frac{7}{2}$

$$D(X)=E\left(X^2\right)-[E(X)]^2=\frac{1}{2}$$

2. 把 4 个球随机地投入 4 个盒子中，设 X 表示空盒子的个数，求：$E(X)$ 和 $D(X)$.

解 先求 X 的概率分布. X 的可能取值为 0, 1, 2, 3. 于是

$$P\{X=0\}=\frac{4!}{4^4}=\frac{6}{64}$$

$$P\{X=1\}=\frac{3C_4^1C_4^1C_3^1}{4^4}=\frac{36}{64}$$

$$P\{X=2\}=\frac{C_4^2(2C_4^3+C_4^2)}{4^4}=\frac{21}{64}$$

$$P\{X=3\}=\frac{4}{4^4}=\frac{1}{64}$$

于是

$$E(X)=0\cdot\frac{6}{64}+1\cdot\frac{36}{64}+2\cdot\frac{21}{64}+3\cdot\frac{1}{64}=\frac{81}{64}$$

$$E(X^2)=0^2\cdot\frac{6}{64}+1^2\cdot\frac{36}{64}+2^2\cdot\frac{21}{64}+3^2\cdot\frac{1}{64}=\frac{129}{64}$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{129}{64}-(\frac{81}{64})^2=\frac{1695}{64^2}$$

3. 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}2(1-x) & 0<x<1\\ 0 & \text{其它}\end{cases}$$

求：$E(X)$ 和 $D(X)$.

解 $E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^1 x\cdot 2(1-x)dx$

$$=\int_0^1 2xdx-\int_0^1 2x^2dx=\frac{1}{3}$$

$D(X)=\int_0^1(x-\frac{1}{3})^2\cdot 2(1-x)dx$

$=2\int_0^1(x^2-\frac{2}{3}x+\frac{1}{9})\cdot(1-x)dx=\frac{1}{18}$

4. 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}1+x & -1\le x\le 0\\ 1-x & 0<x\le 1\\ 0 & \text{其它}\end{cases}$$

求：$E(X)$ 和 $D(X)$.

解 $E(X)=\int_{-1}^{0}x(1+x)dx+\int_0^1 x(1-x)dx=0$

$E(X^2)=\int_{-1}^{0}x^2(1+x)dx+\int_0^1 x^2(1-x)dx=\frac{1}{6}$

于是

$$D(X)=E(X^2)-[E(X)]^2=\frac{1}{6}$$

5. 设 X 表示 10 次独立重复射击命中目标的次数，每次射中目标的概率为 0.4，求 $E(X^2)$.

解 由于 X 服从二项分布，所以 $E(X)=4$ 和 $D(X)=0.24$.

又由于 $E(X^2)=D(X)+[E(X)]^2=16.24$.

6. 已知随机变量 X 服从参数为 2 的泊松分布，求 $E(3X-2)$.

解 因为 X 服从参数为 2 的泊松分布，所以 $E(X)=2$，从而

$$E(3X-2)=3E(X)-2=3\cdot 2-2=4.$$

7. 设一部机器在一天内发生故障的概率为 0.2，机器发生故障时全天停止工作，一周 5 个工作日，若无故障，可获利润 10 万元；发生一次故障仍可获利润 5 万元；若发生两次故障，或利润 0 元；若发生 3 次或 3 次以上故障就要亏损 2 万元. 求一周内的利润期望.

解 设这部机器内有 X 天发生故障，一周的利润为 Y 万元，由题意可知 $X\sim B(5,0.2)$，且

$$Y=\begin{cases}10 & X=0\\ 5 & X=1\\ 0 & X=2\\ -2 & X\ge 3\end{cases}$$

则

$$E(Y)=10\cdot P\{X=0\}+5\cdot P\{X=1\}+0\cdot P\{X=2\}+(-2)\cdot P\{X\ge 3\}$$

$$=10C_5^0(0.2)^0(0.8)^5+5C_5^1(0.2)^1(0.8)^4-2[1-C_5^0(0.2)^0(0.8)^5-C_5^1(0.2)^1(0.8)^4-C_5^2(0.2)^2(0.8)^3]$$

$$=5.20896$$

8. 设某工厂生产的圆盘，其直径在区间 (a,b) 上服从均匀分布，求该圆盘面积的数学期望.

解 设 X 表示圆盘的直径，由题意可知 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其它}\end{cases}$$

于是该圆盘面积的数学期望为

$$E(\frac{\pi}{4}X^2)=\frac{\pi}{4}E(X^2)=\frac{\pi}{4}\int_a^b x^2\frac{1}{b-a}dx$$

$$=\frac{\pi}{12}(b^2+ab+a^2)$$

9. 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}e^{-x} & x>0\\ 0 & \text{其它}\end{cases}$$

求：(1) $Y=2X$；(2) $Y=e^{-2X}$ 的数学期望.

解 (1) 由于 X 服从参数为 1 的指数分布，故 $E(X)=1$.

从而 $E(Y)=E(2X)=2E(X)=2$.

(2) $E(Y)=E(e^{-2X})=\int_0^{+\infty}e^{-2x}e^{-x}dx=\frac{1}{3}$

10. 设随机变量 ξ 和 η 是相互独立的，且服从同一分布，已知 ξ 的分布律为 $P\{\xi=i\}=\frac{1}{3},i=1,2,3$. 又设 $X=\max(\xi,\eta)$，$X=\min(\xi,\eta)$.

(1) 求二维随机变量 (X, Y) 的分布律;

(2) 求 $E(X)$ 和 $E(X/Y)$.

解 (1) (X, Y) 的分布律为

Y \ X	1	2	3
1	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$
2	0	$\frac{1}{9}$	$\frac{2}{9}$
3	0	0	$\frac{1}{9}$

(2) 由(X,Y)的分布律可得关于 X 的边缘分布律为

X	1	2	3
p	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{5}{9}$

故

$$E(X)=1\times\frac{1}{9}+2\times\frac{1}{3}+3\times\frac{5}{9}=\frac{22}{9}$$

$$E(X/Y)=\frac{1}{1}\times\frac{1}{9}+\frac{2}{1}\times\frac{2}{9}+\frac{3}{1}\times\frac{2}{9}+\frac{2}{2}\times\frac{1}{9}+\frac{3}{2}\times\frac{2}{9}+\frac{3}{3}\times\frac{1}{9}=\frac{16}{9}$$

11. 设随机变量(X,Y)的概率密度为

$$f(x,y)=\frac{1}{8}(x+y),\qquad 0<x<2,0<y<2$$

求：$E(X)$，$E(Y)$，$E(XY)$和$E(X^2+Y^2)$.

解 $E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\int_0^2 dx\int_0^2 x\cdot\frac{1}{8}(x+y)dy=\frac{7}{6}$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy=\int_0^2 dx\int_0^2 y\cdot\frac{1}{8}(x+y)dy=\frac{7}{6}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\int_0^2 dx\int_0^2 xy\cdot\frac{1}{8}(x+y)dy=\frac{4}{3}$$

$$E(X^2+Y^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x^2+y^2)f(x,y)dxdy=\int_0^2 dx\int_0^2(x^2+y^2)\cdot\frac{1}{8}(x+y)dy=\frac{10}{3}$$

12. 设随机变量 X,Y 分别服从参数为 2 和 4 的指数分布,

(1) 求：$E(X+Y)$，$E(2X-3Y^2)$.

(2) 设 X,Y 相互独立，求 $E(XY)$，$D(X+Y)$.

解 (1) 由于 X,Y 分别服从参数为 2 和 4 的指数分布，故 $E(X)=\frac{1}{2},E(Y)=\frac{1}{4}$,

$D(Y)=\frac{1}{16}$.

因此

$$E(X+Y)=E(X)+E(Y)=\frac{3}{4},$$

又 $E(Y^2)=D(Y)+[E(Y)]^2=\frac{1}{16}+\frac{1}{16}=\frac{1}{8}$

从而

$$E(2X-3Y^2)=2E(X)-3E(Y^2)=1-3\times\frac{1}{8}=\frac{5}{8}$$

13. 设 $X\sim N(1,2)$, $Y\sim N(0,1)$, 且 X 和 Y 相互独立, 求随机变量 $Z=2X-Y+3$ 的概率密度.

解 因为 $X\sim N(1,2)$, $Y\sim N(0,1)$, 且 X 和 Y 相互独立, 于是

$$E(Z)=E(2X-Y+3)=2E(X)-E(Y)+3=5,$$

$$D(Z)=D(2X-Y+3)=4D(X)+D(Y)=9$$

即有

$$Z=2X-Y+3\sim N(5,9)$$

从而随机变量 $Z=2X-Y+3$ 的概率密度为

$$f(z)=\frac{1}{\sqrt{2\pi}\cdot 3}e^{-\frac{(z-5)^2}{2\cdot 9}}=\frac{1}{3\sqrt{2\pi}}e^{-\frac{(z-5)^2}{18}}$$

14. 设有 10 个猎人正等着野鸭飞过来, 当一群野鸭飞过头顶时, 他们同时开了枪, 但他们每个人都是随机地, 彼此独立地选择自己的目标. 如果每个猎人独立地射中其目标的概率均为 p, 试求当 10 只野鸭飞来时, 没有被击中而飞走的野鸭数的期望值.

解 设

$$X_i=\begin{cases}1 & 第i个野鸭未被击中\\ 0 & 第i个野鸭被击中\end{cases}\quad i=1,2,\cdots 10$$

飞走的野鸭的期望值可表示为

$$E(X_1+X_2+\cdots+X_{10})=E(X_1)+E(X_2)+\cdots+E(X_{10})$$

又由于

$$E(X_i)=P\{X_i=1\}=(1-\frac{p}{10})^{10}$$

因此

$$E(X)=E(X_1+X_2+\cdots+X_{10})=10(1-\frac{p}{10})^{10}$$

15. 一个骰子掷 10 次, 求得到的总点数的期望.

解 令 $X_i(i=1,2,\cdots 10)$ 表示第 i 次掷骰子的点数, 于是总点数的期望可表示为

$$E(X_1+X_2+\cdots+X_{10})=E(X_1)+E(X_2)+\cdots+E(X_{10})$$

又 $E(X_i)=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{7}{2}$

因此

$$E(X)=E(X_1+X_2+\cdots+X_{10})=10\times\frac{7}{2}=35$$

16. 设随机变量 X 和 Y 的联合概率密度为

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

求：$E(X)$，$E(Y)$，$Cov(X,Y)$.

解 关于 X 和 Y 的边缘分布律为

X	-1	0	1
p	0.15	0.5	0.35

Y	0	1
p	0.4	0.6

所以

$$E(X)=(-1)\times 0.15+0\times 0.5+1\times 0.35=0.2$$

$$E(Y)=0\times 0.4+1\times 0.6=0.6$$

又

$$E(XY)=(-1)\times 0\times 0.07+(-1)\times 1\times 0.08+0\times 0\times 0.18$$
$$+0\times 1\times 0.32+1\times 0\times 0.15+1\times 1\times 0.20=0.12$$

因此

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0$$

17. 设随机变量 (X, Y) 的概率密度为

$$f(x,y)=\begin{cases}1 & |y|<x, 0<x<1\\ 0 & \text{其它}\end{cases}$$

求：$E(X)$，$E(Y)$，$Cov(X,Y)$.

解 $E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\int_0^1 dx\int_{-x}^{x}xdy=\int_0^1 2x^2dx=\frac{2}{3}$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy=\int_0^1 dx\int_{-x}^{x}ydy=0$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\int_0^1 dx\int_{-x}^{x}xydy=0$$

故

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0$$

18. 设随机变量服从拉普拉斯分布，其概率密度为

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

(1) 求：$E(X)$ 和 $D(X)$.

(2) 求：X 与 $|X|$ 的协方差，并问 X 与 $|X|$ 是否不相关？

(3) 问 X 与 $|X|$ 是否相互独立？

解 (1) $E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{-\infty}^{+\infty}x\cdot\frac{1}{2}e^{-|x|}dx=0$

而

$$E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)dx=\int_{-\infty}^{+\infty}x^2\cdot\frac{1}{2}e^{-|x|}dx=\int_0^{+\infty}x^2e^{-x}dx=2$$

所以

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

(2)

$$Cov(X,|X|)=E(X|X|)-E(X)E(|X|)$$

$$=\int_{-\infty}^{+\infty}x|x|f(x)dx-0=\int_{-\infty}^{+\infty}x\cdot|x|\cdot\frac{1}{2}e^{-|x|}dx=0$$

故 X 与 $|X|$ 不相关.

(3)

$$P(X\le 1)=\int_{-\infty}^{1}f(x)dx=\int_{-\infty}^{1}\frac{1}{2}e^{-|x|}dx=1-\int_1^{+\infty}\frac{1}{2}e^{-|x|}dx<1$$

又

$$P(|X|\le 1)=\int_{-1}^{1}f(x)dx=\int_{-1}^{1}\frac{1}{2}e^{-|x|}dx=\int_0^1 e^{-x}dx>0$$

故

$$P(X\le 1,|X|\le 1)=P(|X|\le 1)\ne P(X\le 1)P(|X|\le 1)$$

可见 X 与 $|X|$ 不相互独立.

19. 已知随机变量 $X\sim N(1,9)$，$Y\sim N(0,16)$，且 X 和 Y 的相关系数为 $\rho_{XY}=-\frac{1}{2}$.

设 $Z=\frac{X}{3}+\frac{Y}{2}$.

(1) 求 $E(Z)$ 和 $D(Z)$；

(2) 求 X 与 Z 的相关系数.

解　由题意知，$E(X)=1, D(X)=9, E(Y)=0, D(Y)=16$.

而 $Cov(X,Y)=\rho_{XY}\cdot\sqrt{D(X)}\cdot\sqrt{D(Y)}=\left(-\frac{1}{2}\right)\cdot 3\cdot 4=-6$

所以

$$E(Z)=E(\frac{X}{3}+\frac{Y}{2})=\frac{1}{3}E(X)+\frac{1}{2}E(Y)=\frac{1}{3}$$

$$D(Z)=D(\frac{X}{3}+\frac{Y}{2})=D(\frac{X}{3})+D(\frac{Y}{2})+2Cov(\frac{X}{3},\frac{Y}{2})$$

$$=\frac{1}{9}D(X)+\frac{1}{4}D(Y)+\frac{1}{3}Cov(X,Y)$$

$$=\frac{1}{9}\cdot 9+\frac{1}{4}\cdot 16+\frac{1}{3}\cdot(-6)=3$$

(2)　$$Cov(X,Z)=Cov(X,\frac{X}{3}+\frac{Y}{2})$$

$$=\frac{1}{3}Cov(X,X)+\frac{1}{2}Cov(X,Y)$$

$$=\frac{1}{3}\cdot 9+\frac{1}{2}\cdot(-6)=0$$

习题 B

1. 已知随机变量 X 服从二项分布，且 $E(X)=2.4$ 和 $D(X)=1.44$，求二项分布的参数 n,p 的值.

解　由 $E(X)=2.4$，可得 $np=2.4$. 由 $D(X)=1.44$，可得 $np(1-p)=1.44$.

从而由上解得

$$n=6, p=0.4$$

2. 某流水生产线上每个产品不合格的概率为 $p(0<p<1)$，各产品合格与否相互独立，当出现一个不合格品时即停机检修. 设开机后第一次停机时已产生了的产品个数为 X，求 $E(X)$ 和 $D(X)$.

解　记 $A_k=\{生产的第k个产品是合格品\}, k=1,2,\cdots$. 而 X 可能取的值为全体自然数. 由题意得

$$P\{X=k\}=P\left(A_1A_2\cdots A_{k-1}\bar{A}_k\right)$$

$$=P(A_1)P(A_2)\cdots P(A_{k-1})P(\overline{A}_k)=(1-p)^{k-1}p, k=1,2,\cdots$$

于是

$$E(X)=\sum_{k=1}^{\infty}kP\{X=k\}=\sum_{k=1}^{\infty}k(1-p)^{k-1}p$$

因为 $\sum_{k=1}^{+\infty}kx^{k-1}=\sum_{k=1}^{+\infty}(x^k)'=(\sum_{k=1}^{+\infty}x^k)'=\left(\frac{x}{1-x}\right)'=\frac{1}{(1-x)^2}$

所以

$$E(X)=\sum_{k=1}^{\infty}kP\{X=k\}=\sum_{k=1}^{\infty}k(1-p)^{k-1}p=p\cdot\frac{1}{[1-(1-p)]^2}=\frac{1}{p}$$

又因为 $\sum_{k=1}^{+\infty}k^2x^{k-1}=\sum_{k=1}^{+\infty}(kx^k)'=(\sum_{k=1}^{+\infty}(k+1)x^k-\sum_{k=1}^{+\infty}x^k)'=(\sum_{k=1}^{+\infty}(x^{k+1})'-\frac{x}{1-x})'$

$$=(\sum_{k=1}^{+\infty}x^{k+1})''-\left(\frac{x}{1-x}\right)'=(\frac{x^2}{1-x})''-\left(\frac{x}{1-x}\right)'=\frac{1+x}{(1-x)^3}$$

于是

$$E(X^2)=\sum_{k=1}^{\infty}k^2P\{X=k\}=\sum_{k=1}^{\infty}k^2(1-p)^{k-1}p=p\cdot\frac{1+(1-p)}{[1-(1-p)]^3}=\frac{2-p}{p^2}$$

故

$$D(X)=E(X^2)-[E(X)]^2=\frac{2-p}{p^2}-\frac{1}{p^2}=\frac{1-p}{p^2}$$

3. 设随机变量 X 在区间 $(-1,1)$ 上服从均匀分布，随机变量

$$Y=\begin{cases}-1 & X<0\\ 0 & X=0\\ 1 & X>0\end{cases}$$

求：$E(Y)$ 和 $D(Y)$.

解 由题意，X 的概率密度为

$$f(x)=\begin{cases}\frac{1}{2} & -1<x<1\\ 0 & \text{其它}\end{cases}$$

则

$$P(X>0)=\int_0^{+\infty}f(x)dx=\int_0^1\frac{1}{2}dx=\frac{1}{2}$$

$$P(X<0)=\int_{-\infty}^{0} f(x)dx=\int_{-1}^{0}\frac{1}{2}dx=\frac{1}{2}$$

故

$$E(Y)=1\cdot P(X>0)+0\cdot P(X=0)+(-1)\cdot P(X<0)$$

$$=\frac{1}{2}-\frac{1}{2}=0$$

$$E(Y^2)=1^2\cdot P(X>0)+0^2\cdot P(X=0)+(-1)^2\cdot P(X<0)$$

$$=\frac{1}{2}+\frac{1}{2}=1$$

故

$$D(Y)=E(Y^2)-[E(Y)]^2=1$$

4. 设随机变量 X 概率密度为

$$f(x)=\begin{cases}\frac{1}{2}\cos\frac{x}{2} & 0<x<\pi\\ 0 & 其它\end{cases}$$

对 X 独立地观察 4 次，用 Y 表示观察值大于 $\frac{\pi}{3}$ 的次数，求 Y^2 的数学期望.

解　因为 $P\left\{X>\frac{\pi}{3}\right\}=\int_{\frac{\pi}{3}}^{+\infty} f(x)dx=\int_{\frac{\pi}{3}}^{\pi}\frac{1}{2}\cos\frac{x}{2}dx=\frac{1}{2}$

故 $Y\sim B(4,\frac{1}{2})$，得 $E(Y)=4\cdot\frac{1}{2}=2, D(Y)=4\cdot\frac{1}{2}\cdot\frac{1}{2}=1$

所以 $E(Y^2)=D(Y)+[E(Y)]^2=1+4=5$.

5. 设随机变量 Y 服从参数为 1 的指数分布，随机变量

$$X_k=\begin{cases}1 & Y>k\\ 0 & Y\le k\end{cases}\qquad (k=1,2)$$

求：(1) (X_1,X_2) 的分布律；

(2)　$E(X_1+X_2)$.

解　由已知，Y 的概率密度为

$$f(y)=\begin{cases}e^{-y} & y>0\\ 0 & y\le 0\end{cases}$$

(X_1,X_2) 所有可能取值为 (0, 0)，(0, 1)，(1, 0)，(1, 1).

(1) $P(X_1=0,X_2=0)=P(Y\le 1,Y\le 2)=P(Y\le 1)=\int_0^1 e^{-y}dy=1-e^{-1}$

$$P\left(X_1=0,X_2=1\right)=P\left(Y\le 1,Y>2\right)=0$$

$$P\left(X_1=1,X_2=0\right)=P\left(Y>1,Y\le 2\right)=P\left(1<Y\le 2\right)=\int_1^2 e^{-y}dy=e^{-1}-e^{-2}$$

$$P\left(X_1=1,X_2=1\right)=P\left(Y>1,Y>2\right)=P\left(Y>2\right)=\int_2^{+\infty} e^{-y}dy=e^{-2}$$

(2) $E\left(X_1+X_2\right)=E\left(X_1\right)+E\left(X_2\right)=P(Y>1)+P(Y>2)$

$$=\int_1^{+\infty} e^{-y}dy+\int_2^{+\infty} e^{-y}dy=e^{-1}+e^{-2}$$

6. 设 X 和 Y 是两个相互独立且均服从正态分布 $N(0,\frac{1}{2})$ 的随机变量，求 $E\left[|X-Y|\right]$.

解 记 $\xi=X-Y$，由 $X\sim N(0,\frac{1}{2}),Y\sim N(0,\frac{1}{2})$，知

$E(\xi)=E(X)-E(Y)=0$

$D(\xi)=D(X)+D(Y)=\frac{1}{2}+\frac{1}{2}=1$

即

$$\xi\sim N(0,1)$$

所以

$$E\left[|X-Y|\right]=E(|\xi|)=\int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{2}{\sqrt{2\pi}}\int_0^{+\infty}xe^{-\frac{x^2}{2}}dx=\frac{2}{\sqrt{2\pi}}$$

7. 设 A, B 随机事件，且 $P(A)=\frac{1}{4}$，$P(B|A)=\frac{1}{3}$，$P(A|B)=\frac{1}{2}$，令

$$X=\begin{cases}1 & A\text{发生}\\ 0 & A\text{不发生}\end{cases},\quad Y=\begin{cases}1 & B\text{发生}\\ 0 & B\text{不发生}\end{cases}$$

求：(1) 二维随机变量 (X, Y) 的概率密度；(2) X 与 Y 的相关系数.

解 $\because \frac{1}{3}=P\left(B|A\right)=\frac{P(AB)}{P(A)}$ $\quad\therefore P(AB)=\frac{1}{3}P(A)=\frac{1}{3}\cdot\frac{1}{4}=\frac{1}{12}$

又 $\because P(A|B)=\frac{P(AB)}{P(B)}=\frac{1}{2}$ $\quad\therefore P(B)=2P(AB)=\frac{1}{6}$

(1) $P(X=1,Y=1)=P(AB)=\frac{1}{12}$

$P(X=0,Y=1)=P(\bar{A}B)=P(B)-P(AB)=\frac{1}{6}-\frac{1}{12}=\frac{1}{12}$

$P(X=1,Y=0)=P(A\bar{B})=P(A)-P(AB)=\frac{1}{4}-\frac{1}{12}=\frac{1}{6}$

$P(X=0,Y=0)=P(\bar{A}\bar{B})=1-P(A\cup B)=1-P(A)-P(B)+P(AB)=\frac{2}{3}$

故(X,Y)的分布律为

$Y \backslash X$	0	1
0	$\frac{2}{3}$	$\frac{1}{12}$
1	$\frac{1}{6}$	$\frac{1}{12}$

(2) 由(1)易得关于X,Y的分布律分别为

X	0	1
p	$\frac{3}{4}$	$\frac{1}{4}$

Y	0	1
p	$\frac{5}{6}$	$\frac{1}{6}$

故

$E(X)=\frac{1}{4},E(X^2)=\frac{1}{4},$

$D(X)=E(X^2)-[E(X)]^2=\frac{1}{4}-(\frac{1}{4})^2=\frac{3}{16}$

$E(Y)=\frac{1}{6},E(Y^2)=\frac{1}{6},$

$D(Y)=E(Y^2)-[E(Y)]^2=\frac{1}{6}-(\frac{1}{6})^2=\frac{5}{36}$

而由(X,Y)的分布律,可知

$E(XY)=1\cdot1\cdot\frac{1}{12}=\frac{1}{12}$

故得

$$\rho_{XY}=\frac{E(XY)-E(X)E(Y)}{\sqrt{D(X)}\sqrt{D(X)}}=\frac{\frac{1}{12}-\frac{1}{4}\cdot\frac{1}{6}}{\sqrt{\frac{3}{16}}\sqrt{\frac{5}{36}}}=\frac{1}{\sqrt{15}}$$

8.将一枚硬币重复掷 n 次,以 X 和 Y 分别表示正面向上和反面向上的次数,求 X 和 Y 的相关系数.

解 因为 $X+Y=n$,所以 $Y=n-X$.

故 $D(Y)=D(n-X)=D(X)$,

$Cov(X,Y)=Cov(X,n-X)=-Cov(X,X)=-D(X)$

所以 X 和 Y 的相关系数为

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{-D(X)}{\sqrt{D(X)}\sqrt{D(X)}}=-1。$$

第 5 章　大数定律和中心极限定理

习题 A

1．设 X 为随机变量， $EX=\mu$， $DX=\sigma^2$，试估计 $P\{|X-\mu|<3\sigma\}$.

解: 由切比雪夫不等式, 有

$$P\{|X-\mu|<3\sigma\}\ge 1-\frac{DX}{(3\sigma)^2}=1-\frac{1}{9}=\frac{8}{9}.$$

2．某路灯管理所有 20000 只路灯，夜晚每盏路灯开的概率为 0.6，设路灯开关是相互独立的，试用切贝雪夫不等式估计夜晚同时开着的路灯数在 11000-13000 盏之间的概率.

解：记 X 为晚上开着的路灯数，则 $X\sim B(20000,0.6)$，因此

$$EX=20000\times 0.6=12000$$

$$DX=20000\times 0.6\times(1-0.6)=4800$$

由切比雪夫不等式有

$$P\{11000<X<13000\}=P\{|X-12000|<1000\}\ge 1-\frac{48000}{1000^2}=0.9952.$$

3．在 n 重伯努利试验中，若已知每次试验中事件 A 出现的概率为 0.75，请利用切贝雪夫不等式估计 n，使 A 出现的概率在 0.74 至 0.76 之间的概率不小于 0.90.

解: 假设

$$X_i=\begin{cases}1, A\text{出现}\\ 0, A\text{不出现}\end{cases},\quad i=1,2,\cdots,n$$

$X=\sum_{i=1}^{n}X_i$, 则 $EX=np$， $DX=npq$, 其中 $p=0.75$, 所以

$$P\{|X-np|<0.01n\}>1-\frac{npq}{(0.01n)^2}\ge 0.9$$

解得 $n\ge 18750$.

4. 某批产品合格率为 0. 6, 任取 10000 件, 其中恰有合格品在 5980 件到 6020 件之间的概率是多少?

解：假设 X 表示任取 10000 件产品中, 合格品的数量, 则

$$X \sim B(10^4, 0.6)$$

即 $EX = 6000, DX = 2400$,

根据中心极限定理, $\dfrac{X-6000}{\sqrt{2400}}$ 近似服从标准正态分布 $N(0,1)$, 则

$$P\{5980 < X < 6020\} = P\{|X-6000| < 20\} = P\left\{\left|\frac{X-6000}{\sqrt{2400}}\right| < \frac{20}{\sqrt{2400}}\right\}$$

$$\approx 2\Phi\left(\frac{20}{\sqrt{2400}}\right) - 1 = 0.3182$$

5. 某保险公司有 3000 个同一年龄段的人参加人寿保险，在一年中这些人的死亡率为 0.1%. 参加保险的人在一年的开始交付保险费 100 元，死亡时家属可从保险公司领取 10000 元. 求:

(1) 保险公司一年获利不少于 240000 元的概率;

(2) 保险公司亏本的概率.

解: 假设 X 表示一年内死亡的人数, 则

$$X \sim B(3000, 0.001)$$

且 $EX = 3, DX = 2.997$, 并根据中心极限定理, $\dfrac{X-3}{\sqrt{2.997}}$ 近似服从标准正态分布 $N(0,1)$, 则

(1) 保险公司一年内获利不少于 240000 元的概率为:

$$P\{3\times10^5 - 10^4 \times X > 2.4\times10^5\} = P\{X < 6\} = \Phi\left(\frac{3}{\sqrt{2.997}}\right) \approx 0.958.$$

(2) 保险公司亏本的概率为:

$$P\{3\times10^5 - 10^4 \times X < 0\} = P\{X > 30\} = 1 - \Phi\left(\frac{27}{\sqrt{2.997}}\right) \approx 0.$$

6. 计算器在进行加法时，将每个加数舍入最靠近它的整数，设所有舍入误差相互独立且在 $(-0.5, 0.5)$ 上服从均匀分布，

(1) 将 1500 个数相加，问误差总和的绝对值超过 15 的概率是多少?

(2) 最多可有几个数相加使得误差总和的绝对值小于 10 的概率不小于 0.90?

解：假设 X_i 表示每次计算时,所得到的误差,则

$$X_i \sim U(-0.5, 0.5),\ i = 1, 2, \cdots, 1500,$$

$X = \sum_{i=1}^{1500} X_i$ 表示 1500 个数相加,所得到误差总和，$EX = 0, DX = \frac{1500}{12} = 125$，根据中心极限定理，$X/\sqrt{125}$ 近似服从标准正态分布，

(1) $P\{|X| > 15\} = 1 - P\{-15 < X < 15\} \approx 2 - 2\Phi\left(\frac{15}{\sqrt{25}}\right) = 2(1 - 0.9099) = 0.1802$

(2)假设最多可有 n 个数相加使得误差总和的绝对值小于 10 的概率不小于 0.90:

$$P\left\{\left|\sum_{i=1}^{n} X_i\right| < 10\right\} > 0.90 \Rightarrow P\left\{-10 < \sum_{i=1}^{n} X_i < 10\right\} = P\left\{\frac{-10}{\sqrt{n/12}} < \frac{\sum_{i=1}^{n} X_i}{\sqrt{n/12}} < \frac{10}{\sqrt{n/12}}\right\}$$

$$= 2\Phi\left(\frac{10}{\sqrt{n/12}}\right) - 1 > 0.9$$

解得 $n = 443$.

7. 对敌人的防御地带进行 100 次轰炸，每次轰炸命中目标的炸弹数目是一个均值为 2，方差为 1.69 的随机变量．求在 100 次轰炸中有 180 到 220 颗炸弹命中目标的概率.

解：假设 $X_i = \begin{cases} 1, \text{第i次击中目标} \\ 0, \text{第i次没有击中目标} \end{cases}$，$i = 1, 2, \cdots, 100$

$X = \sum_{i=1}^{100} X_i$ 表示 100 次轰炸中,击中目标的总次数,则 $EX = 200, DX = 169$，根据中心极限定理，$\frac{X - 200}{\sqrt{169}}$ 近似服从正态分布,则有

$$P\{180 < X < 220\} = P\left\{\frac{-20}{\sqrt{169}} < \frac{X - 200}{\sqrt{169}} < \frac{20}{\sqrt{169}}\right\}$$

$$= 2\Phi\left(\frac{20}{13}\right) - 1 = 2 \times 0.9382 - 1 = 0.8764$$

8. 有一批建筑房屋用的木柱，其中 80%的长度不小于 3 米，现从这批木柱中随机地取 100 根，求其中至少有 30 根短于 3 米的概率.

解: $X_i=\begin{cases}1,\text{第i根木柱短语3米}\\0,\text{第i根木柱长于3米}\end{cases}$, $i=1,2,\cdots,100$

$X=\sum_{i=1}^{100}X_i$ 表示 100 根木柱中，短于 3 米的数目，则 $X\sim B(100,0.2)$, $EX=20, DX=16$,

$$P\{X>30\}=P\left\{\frac{X-20}{4}>\frac{10}{4}\right\}=1-\Phi(2.5)=1-0.99379=0.00621.$$

习题 B

1. 设 $X_1,X_2,\cdots,X_n$ 是相互独立，服从同一分布的连续型随机变量，且 $EX_i=0$，$DX_i=1$，证明：对任意正数 $\lambda>0$，有 $P\left\{\frac{1}{n}\sum_{i=1}^{n}X_i^2\geq\lambda\right\}\leq\frac{1}{\lambda}$.

证明: 根据题意知，$EX_i^2=1, DX_i^2<\infty$，$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=1$, 则根据切比雪夫不等式, 有

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}X_i^2-E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)\geq\lambda-1\right\}\leq\frac{D\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)}{(\lambda-1)^2}=\frac{DX_i^2}{n(\lambda-1)^2}<\frac{1}{\lambda}$$

上面式子中, 当 n 足够大时, 对任意正数 λ, 最后一个不等式仍成立.

2. 设 $X_1,X_2,\cdots$ 是独立同分布的随机变量列，且 $E\left(|X_1|^k\right)<\infty$，$k$ 为正整数，则 $\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{P}E(X_1^k)$.

证明：$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^k\right)=EX_1^k$, 根据切比雪夫不等式, 有

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i^k-EX_1^k\right|<\varepsilon\right\}\geq1-\frac{D\left(\frac{1}{n}\sum_{i=1}^{n}X_i^k\right)}{\varepsilon^2}$$

当 $n\to\infty$ 时， $DX\to 0$，所以有 $\lim\limits_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i^k-EX_1^k\right|<\varepsilon\right\}=1$，即

$$\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{P}E(X_1^k)$$

成立.

3．分别用切贝雪夫不等式与德莫弗-拉普拉斯定理确定：当掷一枚硬币时，需要掷多少次才能保证出现正面的概率在 0.4 和 0.6 之间的概率不少于 0.9?

解: $X_i=\begin{cases}1，出现正面\\0，出现反面\end{cases}$，$i=1,2,\cdots,n$

$X=\sum\limits_{i=1}^{n}X_i$ 表示掷 n 次硬币，正面向上的次数，则 $E\overline{X}=p=0.5,D\overline{X}=\frac{0.25}{n}$，$\overline{X}=\frac{1}{n}X$，下面分别用切比雪夫不等式和德莫弗-拉普拉斯定理求解 n,

(1) 由切比雪夫不等式:

$$P\left\{0.4n<\sum_{i=1}^{n}X_i<0.6n\right\}=P\left\{\left|\overline{X}-0.5\right|<0.1\right\}\ge 1-\frac{D\overline{X}}{0.1^2}=0.9$$

$$\Rightarrow\frac{0.25}{n}=0.1\Rightarrow n=250$$

(2) 由德莫弗-拉普拉斯定理

$$P\left\{0.4n<\sum_{i=1}^{n}X_i<0.6n\right\}=P\left\{\left|\overline{X}-0.5\right|<0.1\right\}$$

$$=P\left\{\left|\frac{\overline{X}-0.5}{\sqrt{0.25/n}}\right|<\frac{0.1}{\sqrt{0.25/n}}\right\}=2\Phi\left(\frac{0.1}{\sqrt{0.25/n}}\right)-1=0.9$$

$\Rightarrow n>67.24$，即 n 至少要取 68.

4．已知在某十字路口，一周内事故发生数的数学期望为 2.2，标准差为 1.4，

(1) 以 $\overline{X}$ 表示一年内(52 周计)此十字路口事故发生数的算术平均，使用中心极限定理求 $\overline{X}$ 的近似分布，并求 $P\left\{\overline{X}<2\right\}$;

(2) 求一年内事故发生数小于 100 的概率.

解:

(1) 经计算 $E\bar{X}=2.2, D\bar{X}=\frac{1.4^2}{52}$，根据中心极限定理， $\bar{X}$ 近似服从期望为 2.2，方差为 $\frac{1.4^2}{52}$ 的正态分布，即 $\bar{X}\sim N(2.2,1.4^2/52)$. 且

$$P\{\bar{X}<2\}=P\left\{\frac{\bar{X}-2.2}{1.4/\sqrt{52}}<\frac{2-2.2}{1.4/\sqrt{52}}\right\}=\Phi(-\frac{\sqrt{52}}{7})=1-\Phi(\frac{\sqrt{52}}{7})=1-\Phi(1.03)=0.1515$$

(2) 一年内事故发生数少于 100 的概率为:

$$P\left\{\sum_{i=1}^{52}X_i\ge 100\right\}=P\left\{\frac{\frac{1}{52}\sum_{i=1}^{52}X_i-2.2}{1.4/\sqrt{52}}\ge\frac{100/52-2.2}{1.4/\sqrt{52}}\right\}=1-\Phi\left(\frac{100/52-2.2}{1.4/\sqrt{52}}\right)=0.0764$$

5．为检验一种新药对某种疾病的治愈率为 80%是否可靠，给 10 个患该疾病的病人同时服药，结果治愈人数不超过 5 人，试判断该药的治愈率为 80%是否可靠.

解: 假设 $X_i=\begin{cases}1, & \text{第i个人被治愈}\\ 0, & \text{否则}\end{cases}$, $i=1,2,\cdots,n$

则 $P\{X_i=1\}=0.8$, $X=\sum_{i=1}^{10}X_i$ 表示 10 个服用该药的患者的治愈人数，则根据德莫弗-拉普拉斯定理 X 近似服从 $N(8,1.6)$，所以

$$P\left\{\sum_{i=1}^{10}X_i<5\right\}=P\left\{\frac{\sum_{i=1}^{10}X_i-8}{\sqrt{1.6}}<\frac{5-8}{\sqrt{1.6}}\right\}=\Phi\left(-\frac{3}{\sqrt{1.6}}\right)=1-\Phi(2.37)=0.0089$$

由此可以看出假定治愈率为 80%是不可靠的.

6．一公寓有 200 户住户，一户住户拥有汽车辆数 X 的分布律为

X	0	1	2
p_k	0.1	0.6	0.3

问需要多少车位，才能使每辆汽车都具有一个车位的概率至少为 0.95?

解: 假设 X_i 表示第 i 户人家拥有的汽车数，则 $EX_i=0\times0.1+1\times0.6+2\times0.3=1.2$，

$DX_i = EX_i^2 - EX_i = 0^2\times 0.1 + 1^2\times 0.6 + 2^2 \times 0.3 - 1.2^2 = 0.36$，根据中心极限定理，

$\sum_{i=1}^{200} X_i$ 近似服从 $N(1.2\times 200, 0.36\times 200)$，所以假设需要 n 个车位，才能使每辆汽车都具有一个车位的概率至少为 0.95，即

$$P\left\{\sum_{i=1}^{200} X_i \le n\right\} = P\left\{\frac{\sum_{i=1}^{200} X_i - 240}{\sqrt{72}} \le \frac{n-240}{\sqrt{72}}\right\} \ge 0.95$$

$$\Rightarrow \Phi\left(\frac{n-240}{\sqrt{72}}\right) = 0.95 \Rightarrow \frac{n-240}{\sqrt{72}} = 1.64 \Rightarrow n = 254.$$

7. 甲、乙两个戏院在竞争 1000 名观众，假设每个观众可随意选择戏院，观众之间相互独立，问每个戏院应该设有多少座位才能保证因缺少座位而使观众离去的概率小于 1%.

解：假设 $X_i = \begin{cases} 1, & \text{第i名观众选择甲戏院} \\ 0, & \text{第i名观众选择乙戏院} \end{cases}, i = 1, 2, \cdots, 1000$

$X = \sum_{i=1}^{1000} X_i$ 表示 1000 名观众中选择甲戏院的人数，则根据已知

$$P\{X_i = 1\} = P\{X_i = 0\} = p = 0.5,$$

则

$$EX = np = 500, DX = npq = 250$$

根据德莫弗−拉普拉斯定理，X 近似服从 $N(500, 250)$. 由假设每个戏院设有 n 个座位才能保证因缺少座位而使观众离去的概率小于 1%，即

$$P\left\{\sum_{i=1}^{1000} X_i > n\right\} = P\left\{\frac{\sum_{i=1}^{1000} X_i - 500}{\sqrt{250}} > \frac{n-500}{\sqrt{250}}\right\} = 1 - \Phi\left(\frac{n-500}{\sqrt{250}}\right) < 0.01$$

$$\Rightarrow \Phi\left(\frac{n-500}{\sqrt{250}}\right) = 0.9 \Rightarrow \frac{n-500}{\sqrt{250}} = 2.33 \Rightarrow n = 537$$

第 6 章　数理统计的基本概念

习题 A

1．选择题

(1) 由本章定理 3 知 $\bar{X}$ 和 S^2 独立，即（A）成立，从而 $\bar{X}$ 与 S^2 的函数也独立，所以（B）、（C）成立。

(2)

解：$\dfrac{\bar{X}_1-\mu}{\sigma/\sqrt{10}} \sim N(0,1)$，$\dfrac{\bar{X}_2-\mu}{\sigma/\sqrt{15}} \sim N(0,1)$，所以

$$p_1 = P\left\{\left|\bar{X}_1-\mu\right|>\sigma\right\} = P\left\{\frac{\bar{X}_1-\mu}{\sigma/\sqrt{10}}>\sqrt{10}\right\}=1-\Phi(\sqrt{10})$$

$$p_2 = P\left\{\left|\bar{X}_2-\mu\right|>\sigma\right\} = P\left\{\frac{\bar{X}_2-\mu}{\sigma/\sqrt{15}}>\sqrt{15}\right\}=1-\Phi(\sqrt{15})$$

$\Rightarrow p_1 > p_2$，选（C）

(3)

解：$\bar{X} \sim N(1,\dfrac{2^2}{n})$，即 $\dfrac{\bar{X}-1}{2/\sqrt{n}} \sim N(0,1)$，所以选（C）

(4)

解：所谓统计量，即不含有任何未知参数的样本的函数。根据题意，μ 已知，σ^2 未知，所以非统计量的应该是（D）

2．在总体 $N(80,20^2)$ 中随机抽取一容量为 100 的样本，求样本均值与总体均值的差的绝对值大于 3 的概率.

解：假设样本均值为 $\bar{X}$，根据题意，$\bar{X} \sim N(80,\dfrac{20^2}{100})$，则样本均值与总体均值的差的绝对值大于 3 的概率为：

$$P\left\{\left|\bar{X}-80\right|>3\right\} = P\left\{\frac{\left|\bar{X}-80\right|}{2}>\frac{3}{2}\right\}=2(1-\Phi(1.5))=2(1-0.9332)=0.1336.$$

3．已知总体 $X \sim N(20,3)$，$\bar{X}_1$ 和 $\bar{X}_2$ 分别为该总体容量为 10 和 15 的两个样

本均值，而且它们相互独立，试求

$$P\left\{\left|\bar{X}_1-\bar{X}_2\right|>0.3\right\}.$$

解：根据题意，$\bar{X}_1 \sim N(20,\frac{3}{10})$，$\bar{X}_2 \sim N(20,\frac{3}{15})$ 且相互独立，所以 $\bar{X}_1-\bar{X}_2 \sim N(0,\frac{1}{2})$，则

$$P\left\{\left|\bar{X}_1-\bar{X}_2\right|>0.3\right\}=P\left\{\frac{\left|\bar{X}_1-\bar{X}_2\right|}{1/\sqrt{2}}>\frac{0.3}{1/\sqrt{2}}\right\}=2(1-\Phi\left(0.3\times\sqrt{2}\right))=2(1-0.6628)=0.6744$$

4. 在总体 $N(52,6.3^2)$ 中随机抽取一容量为 36 的样本，求样本均值 $\bar{X}$ 落在 50.2 到 53.8 之间的概率.

解: 根据已知条件知，$\bar{X} \sim N(52,\frac{6.3^2}{36})$,

$$P\left\{50.8<\bar{X}<53.8\right\}=P\left\{\frac{-1.2}{6.3/6}<\frac{\bar{X}-52}{6.3/6}<\frac{1.8}{6.3/6}\right\}=\Phi\left(\frac{1.8}{6.3/6}\right)-\Phi\left(\frac{-1.2}{6.3/6}\right)$$

$$=\Phi\left(\frac{1.8}{6.3/6}\right)-1+\Phi\left(\frac{1.2}{6.3/6}\right)=\Phi\left(1.714\right)+\Phi\left(1.143\right)-1$$

$$=0.9564-0.8729-1=0.8293$$

5. 设 $X_1,X_2,\cdots,X_{10}$ 为总体 $N(0,0.3^2)$ 的一个样本，求 $P\left\{\sum_{i=1}^{10}X_i^2>1.44\right\}$.

解: 根据已知条件知，$\frac{\sum_{i=1}^{10}X_i^2}{0.09} \sim \chi^2(10)$

所以 $P\left\{\sum_{i=1}^{10}X_i^2>1.44\right\}=P\left\{\frac{\sum_{i=1}^{10}X_i^2}{0.09}>\frac{1.44}{0.09}\right\}=P\left\{\frac{\sum_{i=1}^{10}X_i^2}{0.09}>16\right\}=0.1$

6. 查表计算：(1) $\chi^2_{0.05}(9)$；(2) $\chi^2_{0.99}(21)$；(3) $t_{0.025}(16)$；(4) $t_{0.01}(56)$；(5) $F_{0.05}(12,20)$；(6) $F_{0.95}(12,20)$.

解: 经查表得: (1) $\chi^2_{0.05}(9)$=16.92, (2) $\chi^2_{0.99}(21)$=8.90, (3) $t_{0.025}(16)$=2.120, (4) $t_{0.01}(56)\approx z_{0.01}=2.33$, (5) $F_{0.05}(12,20)$=2.28,

(6) $F_{0.95}(12,20)=\dfrac{1}{F_{0.05}(20,12)}=0.3937$.

7．设总体 $X\sim N(0,1)$，$X_1,X_2,\cdots,X_n$ 为简单随机样本，试问下列统计量各服从什么分布？

(1) $\dfrac{X_1-X_2}{\sqrt{X_3^2+X_4^2}}$；　　(2) $\dfrac{(\frac{n}{3}-1)\sum\limits_{i=1}^{3}X_i^2}{\sum\limits_{i=4}^{n}X_i^2}$.

解：(1) 根据已知条件知，$X_i\sim N(0,1)$, $i=1,2,\cdots,n$. 所以

$$X_1-X_2\sim N(0,2),\ X_3^2+X_4^2\sim\chi^2(2)\Rightarrow\frac{(X_1-X_2)/\sqrt{2}}{\sqrt{(X_3^2+X_4^2)/2}}\sim t(2)$$

$$\Rightarrow\frac{X_1-X_2}{\sqrt{X_3^2+X_4^2}}\sim t(2)$$

(2) $\sum\limits_{i=1}^{3}X_i^2\sim\chi^2(3)$, $\sum\limits_{i=4}^{n}X_i^2\sim\chi^2(n-3)$, 所以

$$\frac{\sum\limits_{i=1}^{3}X_i^2/3}{\sum\limits_{i=4}^{n}X_i^2/(n-3)}\sim F(3,n-3)\Rightarrow\frac{(\frac{n}{3}-1)\sum\limits_{i=1}^{3}X_i^2}{\sum\limits_{i=4}^{n}X_i^2}\sim F(3,n-3)$$

8. 设 $X\sim N(0,1)$, $Y\sim\chi^2(n)$, X 与 Y 相互独立，又 $t=\dfrac{X}{\sqrt{Y/n}}$，证明：$t^2\sim F(1,n)$.

证明：$t=\dfrac{X}{\sqrt{Y/n}}\Rightarrow t^2=\dfrac{X^2}{Y/n}$，根据已知得，$X^2\sim\chi^2(1)$，所以

$$t^2=\frac{X^2}{Y/n}=\frac{X^2/1}{Y/n}\sim F(1,n)$$

9．$X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim N(\mu,\sigma^2)$ 的样本，$\overline{X}$ 和 S^2 分别是样本均值和样本方差，试求 $E\overline{X}$、$D\overline{X}$ 和 ES^2.

解：$E\overline{X}=E\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i\right)=\dfrac{1}{n}\sum\limits_{i=1}^{n}E(X_i)=\dfrac{1}{n}\sum\limits_{i=1}^{n}\mu=\mu$；

$$D\bar{X}=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{1}{n}\sigma^2;$$

$$ES^2=E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}EX_i^2-nE\bar{X}^2\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}\left(DX_i+(EX_i)^2\right)-n\left((E\bar{X})^2+D\bar{X}\right)\right)=\frac{1}{n-1}\left[\sum_{i=1}^{n}\left(\sigma^2+\mu^2\right)-n\left(\mu^2+\frac{1}{n}\sigma^2\right)\right]=\sigma^2$$

10. 从某总体中抽取一个样本，样本观测值为 2，1，－1，2，试求经验分布数 $F_4(x)$.

解： $F_4(x)=\begin{cases}0, x\le -1\\ \frac{1}{4}, -1<x\le 1\\ \frac{2}{4}=\frac{1}{2}, 1<x\le 2\\ 1, x>2\end{cases}$

习题 B

1. 选择题:

(1) 设随机变量 $X\sim N(0,1)$， $Y\sim N(0,1)$，则下列结论正确的是（　　）

(A) $X+Y$ 服从正态分布;　　(B) X^2+Y^2 服从 χ^2 分布;

(C) X^2/Y^2 服从 F 分布;　　(D) X^2 和 Y^2 均服从 χ^2 分布.

解：由已知 $X\sim N(0,1)$， $Y\sim N(0,1)$，所以有 X^2 和 Y^2 均服从 $\chi^2(1)$ 分布，因此（D）正确。但是已知条件并没有告诉我们， X,Y 是否相互独立，所以（A）、（B）、（C）都不正确。

(2) 设 $X_1,X_2,\cdots,X_n$ 来自总体 $\chi^2(n)$， $\bar{X}$ 表示样本均值，则（　　）.

(A) $E\bar{X}=n, D\bar{X}=2$;　　(B) $E\bar{X}=1, D\bar{X}=2$;

(C) $E\bar{X}=n, D\bar{X}=2n$;　　(D) $E\bar{X}=1, D\bar{X}=2n$.

解：这是由 χ^2 分布的性质知，若 $X\sim\chi^2(n)$，则 $EX=n, DX=2n$，所得答案 (A)

正确。

(3) 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本， $\bar{X}$ 是样本均值，记

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2，\quad S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2，$$

$$S_3^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2，\quad S_4^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 .$$

则服从自由度为 $n-1$ 的 t 分布的随机变量是(　　)，

(A) $t = \dfrac{\bar{X} - \mu}{S_1 / \sqrt{n-1}}$；　　(B) $t = \dfrac{\bar{X} - \mu}{S_2 / \sqrt{n-1}}$；

(C) $t = \dfrac{\bar{X} - \mu}{S_3 / \sqrt{n-1}}$；　　(D) $t = \dfrac{\bar{X} - \mu}{S_4 / \sqrt{n-1}}$.

解：$\dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$，$\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$，且这两个随机变量相互独立，所以 $\dfrac{\dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 / (n-1)}} \sim t(n-1)$，对比得，答案（B）正确。

(4) 设 $X_1, X_2, \cdots, X_8$ 和 $Y_1, Y_2, \cdots, Y_{10}$ 分别是来自两个正态总体 $N(-1, 4)$ 和 $N(2, 5)$ 的样本，且相互独立，S_1^2 和 S_2^2 分别是两个样本的样本方差，则服从 $F(7,9)$ 的统计量是(　　).

(A) $\dfrac{4S_1^2}{5S_2^2}$；　(B) $\dfrac{5S_1^2}{4S_2^2}$；　(C) $\dfrac{4S_2^2}{5S_1^2}$；　(D) $\dfrac{5S_1^2}{2S_2^2}$.

解：$\dfrac{7}{4}S_1^2 \sim \chi^2(7)$，$\dfrac{10}{5}S_1^2 \sim \chi^2(9)$，且他们相互独立，所以有

$\dfrac{\frac{7}{4}S_1^2 / 7}{\frac{9}{5}S_1^2 / 9} \sim F(7,9)$，所以答案(B)正确。

2. 填空题:

(1) 从正态总体 $N(1, \sigma^2)$ 中抽取样本 $X_1, X_2, \cdots, X_{10}$，$\bar{X}$，S^2 分别表示样本均值和样本方差，已知 $P\left\{\bar{X} \le 1, S^2 \le \sigma^2\right\} = \dfrac{1}{3}$，则 $P\{S \le \sigma\} =$ ________.

解:

$$P\left\{\bar{X}\le 1, S^2\le\sigma^2\right\}=P\left\{\bar{X}\le 1\right\}P\left\{S^2\le\sigma^2\right\}=P\left\{\frac{\bar{X}-1}{\sigma}\le 0\right\}P\left\{\frac{9S^2}{\sigma^2}\le 9\right\}=\frac{1}{2}P\left\{S\le\sigma\right\}=\frac{1}{3}$$

所以 $P\left\{S\le\sigma\right\}=\frac{2}{3}$

(2) 设随机变量 X 和 Y 相互独立且都服从正态分布 $N(0,3^2)$，而 $X_1,X_2,\cdots,X_9$ 和 $Y_1,Y_2,\cdots,Y_9$ 分别是来自正态总体 X 和 Y 的简单随机样本，则统计量 $U=\dfrac{3\sum_{i=1}^{9}X_i}{\sqrt{\sum_{i=1}^{9}Y_i^2}}$ 服从________分布，参数为__________.

解：$\dfrac{\sum_{i=1}^{9}X_i}{3}\sim N(0,1)$，$\dfrac{\sum_{i=1}^{9}Y_i^2}{3^2}\sim\chi^2(9)$，且相互独立，所以

$$U=\frac{3\sum_{i=1}^{9}X_i}{\sqrt{\sum_{i=1}^{9}Y_i^2}}=\frac{\dfrac{\sum_{i=1}^{9}X_i}{3}}{\sqrt{\dfrac{\sum_{i=1}^{9}Y_i^2}{3^2}/9}}\sim t(9).$$

(3) 设总体 $X\sim N(0,2^2)$，而 $X_1,X_2,\cdots,X_{15}$ 是来自总体的简单随机样本，则随机变量 $Y=\dfrac{X_1^2+X_2^2+\cdots+X_{10}^2}{2(X_{11}^2+X_{22}^2+\cdots+X_{15}^2)}$ 服从________分布，自由度为___________.

解: $\left(X_1^2+X_2^2+\cdots+X_{10}^2\right)/2^2\sim\chi^2(10)$, $(X_{11}^2+X_{22}^2+\cdots+X_{15}^2)/2^2\sim\chi^2(5)$, 所以

$$Y=\frac{X_1^2+X_2^2+\cdots+X_{10}^2}{2(X_{11}^2+X_{22}^2+\cdots+X_{15}^2)}=\frac{\dfrac{\left(X_1^2+X_2^2+\cdots+X_{10}^2\right)}{4}/10}{\dfrac{(X_{11}^2+X_{22}^2+\cdots+X_{15}^2)}{4}/5}\sim F(10,5)$$

(4) 设 X_1,X_2,X_3,X_4 是来自正态总体 $N(0,2^2)$ 的简单随机样本，$X=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$，则当 $a=$____________，$b=$___________时，统计量 X 服从 χ^2 分布，其自由度为___________.

解：$X_1-2X_2\sim N(0,20)$，$3X_3-4X_4\sim N(0,100)$，所以 $\left(X_1-2X_2\right)^2/20\sim\chi^2(1)$，

$\left(3X_3-4X_4\right)^2/100\sim\chi^2(1)$，且它们相互独立，所以有

$\left(X_1-2X_2\right)^2/20+\left(3X_3-4X_4\right)^2/100\sim\chi^2(2)$，即 $a=\dfrac{1}{20},b=\dfrac{1}{100}$，自由度为 2.

3. 设在总体 $N(\mu,\sigma^2)$ 中抽取一容量为 16 的样本，μ 和 σ^2 均未知.

(1) 求 $P\left\{\dfrac{S^2}{\sigma^2}\le 2.041\right\}$，其中 S^2 为样本方差；

(2) 求 $D(S^2)$.

解：(1) $P\left\{\dfrac{S^2}{\sigma^2}\le 2.041\right\}=P\left\{\dfrac{15S^2}{\sigma^2}\le 15\times 2.041\right\}=0.99$（已知 $\dfrac{15S^2}{\sigma^2}\sim\chi^2(15)$，然后经查表得以上概率为 0.99）；

(2) 由 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，所以 $D\left(\dfrac{(n-1)S^2}{\sigma^2}\right)=2(n-1)$

$\Rightarrow\dfrac{(n-1)^2}{\sigma^4}DS^2=2(n-1)$，$\Rightarrow DS^2=\dfrac{2\sigma^4}{15}$

4. (1) 设样本 $X_1,X_2,\cdots,X_6$ 来自总体 $N(0,1)$，$Y=(X_1+X_2+X_3)^2+(X_4+X_5+X_6)^2$，试确定常数 C 使得 CY 服从 χ^2 分布；

(2) 设样本 $X_1,X_2,\cdots,X_5$ 来自总体 $N(0,1)$，$Y=\dfrac{C(X_1+X_2)}{\sqrt{X_3^2+X_4^2+X_5^2}}$，试确定常数 C 使 Y 服从 t 分布.

解：(1) $X_1+X_2+X_3\sim N(0,3)$，$X_4+X_5+X_6\sim N(0,3)$，即

$\dfrac{X_1+X_2+X_3}{\sqrt{3}}\sim N(0,1)$，$\dfrac{X_4+X_5+X_6}{\sqrt{3}}\sim N(0,1)$，则

$\dfrac{1}{3}(X_1+X_2+X_3)^2+\dfrac{1}{3}(X_4+X_5+X_6)^2=\dfrac{1}{3}Y\sim\chi^2(2)$，即 $C=\dfrac{1}{3}$

(2) $\left(X_1+X_2\right)/\sqrt{2}\sim N(0,1)$，$\left(X_3^2+X_4^2+X_5^2\right)\sim\chi^2(3)$，且它们相互独立，所以

$$\frac{(X_1+X_2)/\sqrt{2}}{\sqrt{(X_3^2+X_4^2+X_5^2)/3}} \sim t(3)\text{，所以}C=\sqrt{3/2}$$

5．设样本 $X_1,X_2,\cdots,X_n,X_{n+1}$ 来自总体 $X\sim N(\mu,\sigma^2)$，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，试证统计量:

$$t=\frac{X_{n+1}-\bar{X}}{S}\sqrt{\frac{n}{n+1}}\sim t(n-1).$$

证明: $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$，$X_{n+1},\bar{X}$ 相互独立，所以 $X_{n+1}-\bar{X}\sim N(0,\frac{n+1}{n}\sigma^2)$，又有

$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，且 $X_{n+1}-\bar{X}$ 与 S^2 相互独立，所以

$$t=\frac{X_{n+1}-\bar{X}}{S}\sqrt{\frac{n}{n+1}}=\frac{(X_{n+1}-\bar{X})/\sqrt{\frac{n+1}{n}\sigma^2}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}\sim t(n-1).$$

6．设 $\bar{X}$ 和 S_X^2 分别是样本 $X_1,X_2,\cdots,X_n$ 的样本均值和样本方差，做线性变换 $Y_i=bX_i-a$（a,b 为常数，$b\neq 0$），记 $\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$，$S_Y^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$，试证:

(1) $\bar{X}=\frac{1}{b}(\bar{Y}+a)$;

(2) $S_X^2=\frac{1}{b^2}S_Y^2$.

证明: (1) $X_i=\frac{Y_i+a}{b}$，所以

$$X_i=\frac{Y_i+a}{n},\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i+a}{b}=\frac{1}{b}\times\frac{1}{n}\sum_{i=1}^{n}Y_i+a=\frac{1}{b}(\bar{Y}+a);$$

(2)

$$S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\sum_{i=1}^{n}(\frac{Y_i+a}{b}-\frac{1}{b}(\bar{Y}+a))^2=\frac{1}{n-1}\times\frac{1}{b^2}\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\frac{1}{b^2}S_Y^2$$

7．设随机变量 X 服从分布 $F(n,m)$，求 $\frac{1}{X}$ 的分布.

解：设 Y_1,Y_2 相互独立，且分别服从自由度为 n,m 的 χ^2 分布，则

$X=\dfrac{Y_1/n}{Y_2/m}\sim F(n,m)$，由此有 $\dfrac{1}{X}=\dfrac{Y_2/m}{Y_1/n}\sim F(m,n)$.

第 7 章　参数估计

习题 A

1．填空题:

(1) 设 $X_1,X_2,\cdots,X_n\ (n\ge 3)$ 为总体 $X\sim N(\mu,\sigma^2)$ 的样本，当 $2\overline{X}-X_1$，$\overline{X}$ 及 $\dfrac{1}{5}X_1+\dfrac{3}{10}X_2+\dfrac{1}{2}X_3$ 作为 μ 的无偏估计时，最有效的是___________;

解: $D\overline{X}=\dfrac{1}{n}\sigma^2$;

$D\left(2\overline{X}-X_1\right)=4D\overline{X}+DX_1=4\times\dfrac{1}{n}\sigma^2+\sigma^2=\dfrac{4+n}{n}\sigma^2$;

$D\left(\dfrac{1}{5}X_1+\dfrac{3}{10}X_2+\dfrac{1}{2}X_3\right)=\dfrac{1}{25}DX_1+\dfrac{9}{100}DX_2+\dfrac{1}{4}DX_3=\dfrac{38}{100}\sigma^2$;

又由于 $(n\ge 3)$，经比较可得 $\overline{X}$ 是最有效的.

(2) 已知总体 $X\sim N(\mu,\sigma^2)$，其中 μ 为已知常数. $X_1,X_2,\cdots,X_n$ 是取自总体 X 的一个简单随机样本，如果用统计量 $\hat{\sigma}=\dfrac{c}{n}\sum\limits_{i=1}^{n}|X_i-\mu|$ 作为 σ 的无偏估计，则 $c=$________.

解: $E\hat{\sigma}=\dfrac{c}{n}\sum\limits_{i=1}^{n}|X_i-\mu|=cE|X-\mu|=c\int_{-\infty}^{\infty}|x-\mu|\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$

$=c\int_{-\infty}^{\infty}|t|\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{t^2}{2\sigma^2}}dt=\dfrac{2c}{\sqrt{2\pi}\sigma}\int_0^{\infty}te^{-\frac{t^2}{2\sigma^2}}dt=\dfrac{2c}{\sqrt{2\pi}\sigma}(-\sigma^2)e^{-\frac{t^2}{2\sigma^2}}\Big|_0^{\infty}=\dfrac{\sqrt{2}c\sigma}{\sqrt{\pi}}$

若 $\hat{\sigma}=\dfrac{c}{n}\sum\limits_{i=1}^{n}|X_i-\mu|$ 为 σ 的无偏估计，则有 $E\hat{\sigma}=\sigma$，即

$$\frac{\sqrt{2}c\sigma}{\sqrt{\pi}}=\sigma \Rightarrow c=\sqrt{\frac{\pi}{2}}$$

(3) 在天平上重复称量一重为 a 的物品，假设各次称量结果相互独立且同服从正态分布 $N(a,0.2^2)$. 若以 $\overline{x}$ 表示 n 次称量结果的算术平均值，则为使

$$P\left\{\left|\overline{x}-a\right|<0.1\right\}\geq 0.95,$$

n 的最小值应不小于自然数__________.

解：$P\left\{\left|\overline{x}-a\right|<0.1\right\}=P\left\{\frac{\left|\overline{x}-a\right|}{0.2/\sqrt{n}}<\frac{0.1}{0.2/\sqrt{n}}\right\}=\Phi\left(\frac{\sqrt{n}}{2}\right)\geq 0.95$, 经查表得

$\frac{\sqrt{n}}{2}=1.64 \Rightarrow n=3.24^2=10.2976$, 即 n 的最小值不应小于 11.

2. 选择题

(1) 设 $X_1,X_2,\cdots,X_n$ 是取自总体 X 的一个简单随机样本，则 EX^2 的矩估计量是(　　)

(A) $S_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$; (B) $S_2^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

(C) $S_1^2+\overline{X}^2$; (D) $S_2^2+\overline{X}^2$.

解：$EX^2=DX+(EX)^2$, DX 的矩估计量为 S_2^2, EX 的矩估计量为 $\overline{X}$, 所以 EX^2 的矩估计量是 (D)

(2) 设 $X_1,X_2,\cdots,X_n$ 是取自总体 $X\sim N(0,\sigma^2)$ 的样本，则未知参数 σ^2 的无偏估计量为(　　).

(A) $\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}X_i^2$; (B) $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$;

(C) $\hat{\sigma}^2=\frac{1}{n+1}\sum_{i=1}^{n}X_i^2$; (D) $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$.

解：$E\sum_{i=1}^{n}X_i^2=\sum_{i=1}^{n}EX_i^2=\sum_{i=1}^{n}\left[DX_i+(EX_i)^2\right]=n(\sigma^2+\mu^2)=n\sigma^2$, 不难看出正确答案为 (B).

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=E\left[\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}\right]=\frac{1}{n}\sum_{i=1}^{n}EX_i^2-E\bar{X}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left[DX_i+\left(EX_i\right)^2\right]-\left[\frac{\sigma^2}{n}+\mu^2\right]=\left(\sigma^2+\mu^2\right)-\left[\frac{\sigma^2}{n}+\mu^2\right]=\frac{n-1}{n}\sigma^2\neq\sigma^2$$

(3) 设 $X\sim N(\mu,\sigma^2)$，其中 μ 已知，$\sigma^2\neq 0$ 为未知参数，$X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，样本均值为 $\bar{X}$，则 σ^2 的极大似然估计量为（　　）.

(A) $\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$；　　(B) $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$；

(C) $\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2$；　　(D) $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$.

解：X 的概率密度为

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

似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}},$$

上式两边同时求对数，得对数似然函数为

$$\ln L(\mu,\sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{(x_i-\mu)^2}{2\sigma^2},$$

由于 μ 已知，所以上式两边只需关于 σ^2 求导，并令其等于 0，得对数似然方程

$$\frac{d\ln L}{d\sigma^2}=-\frac{n}{2\sigma^2}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}=0.$$

解上式方程，得极大似然估计值为

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2.$$

所以 σ^2 的极大似然估计量为 (D).

3. 设总体 X 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{\theta-1}, & 1<x<\theta,\\ 0, & \text{其它}.\end{cases}$$

$X_1,X_2,\cdots,X_n$ 是取自总体 X 的简单随机样本.

(1) 求 θ 的矩估计量 $\hat{\theta}$;

(2) 求 $\hat{\theta}$ 的方差 $D\hat{\theta}$.

解: (1) $EX=\int_1^\theta \dfrac{1}{\theta-1}xdx=\dfrac{1}{2(\theta-1)}(\theta^2-1)=\dfrac{\theta+1}{2}$, 所以

$$\hat{EX}=\frac{\hat{\theta}+1}{2}=\bar{X}\Rightarrow\hat{\theta}=2\bar{X}-1;$$

(2) $EX^2=\int_1^\theta \dfrac{1}{\theta-1}x^2dx=\dfrac{1}{3(\theta-1)}\left(\theta^3-1\right)=\dfrac{\theta^2+\theta+1}{3}$

$$DX=EX^2-(EX)^2=\frac{\theta^2+\theta+1}{3}-\left(\frac{\theta+1}{2}\right)^2=\frac{(\theta-1)^2}{12}$$

$$D\hat{\theta}=D(2\bar{X}-1)=4D\bar{X}=4\times\frac{(\theta-1)^2}{12n}=\frac{(\theta-1)^2}{3n}$$

4. 设总体服从正态分布 $N(\mu,1)$, X_1,X_2,X_3 为其样本, 试证下述三个估计量

(1) $\hat{\mu}_1=\dfrac{1}{5}X_1+\dfrac{3}{10}X_2+\dfrac{1}{2}X_3$;

(2) $\hat{\mu}_2=\dfrac{1}{3}X_1+\dfrac{1}{4}X_2+\dfrac{5}{12}X_3$;

(3) $\hat{\mu}_3=\dfrac{1}{3}X_1+\dfrac{1}{6}X_2+\dfrac{1}{2}X_3$

都是 μ 的无偏估计量, 并求出每一个估计量的方差, 问哪一个估计最有效?

证明: (1) $E\hat{\mu}_1=\dfrac{1}{5}EX_1+\dfrac{3}{10}EX_2+\dfrac{1}{2}EX_3=\dfrac{2+3+5}{10}\mu=\mu$;

(2) $E\hat{\mu}_2=\dfrac{1}{3}EX_1+\dfrac{1}{4}EX_2+\dfrac{5}{12}EX_3=\dfrac{4+3+5}{12}\mu=\mu$;

(3) $E\hat{\mu}_3=\dfrac{1}{3}EX_1+\dfrac{1}{6}EX_2+\dfrac{1}{2}EX_3=\dfrac{2+1+3}{6}\mu=\mu$;

所以上述三个估计量都是 μ 的无偏估计;

(1) $D\hat{\mu}_1=\frac{1}{25}DX_1+\frac{9}{100}DX_2+\frac{1}{4}DX_3=\frac{4+9+25}{100}=\frac{38}{100}$;

(2) $D\hat{\mu}_2=\frac{1}{9}DX_1+\frac{1}{16}DX_2+\frac{25}{144}DX_3=\frac{16+9+25}{144}\approx 0.35$;

(3) $D\hat{\mu}_3=\frac{1}{9}DX_1+\frac{1}{36}DX_2+\frac{1}{4}DX_3=\frac{4+1+9}{36}\approx 0.42$

不难看出 $\hat{\mu}_2$ 的方差最小，即 $\hat{\mu}_2$ 最有效.

5. 设总体 X 有分布律

X	-1	0	2
p	2θ	θ	$1-3\theta$

其中 $0<\theta<\frac{1}{3}$ 为待估参数，求 θ 的矩估计.

解: $EX=(-1)\times 2\theta+0\times\theta+2\times(1-3\theta)=2-8\theta$，所以

$$\hat{EX}=2-8\hat{\theta}=\bar{X}\Rightarrow\hat{\theta}=\frac{2-\bar{X}}{8}$$

6. 设一批灯泡的寿命服从正态分布 $N(\mu,\sigma^2)$，现从中抽取一个样本(单位: 小时)的数据如下:

1458，1395，1562，1614，1351，1490，1478，1382，1536，1496

试用矩估计法对这批灯泡的平均使用寿命 μ 及寿命方差 σ^2 作出矩估计.

解: 已知期望和方差的矩估计量分别为 $\bar{X}$ 和 $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$，所以经计算得

$$\hat{\mu}=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=1476.2;$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^{\,2}-\bar{x}^2=6198.56$$

7. 已知某电子设备的使用寿命 X 服从指数分布，其密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & x\ge 0,\\ 0, & x<0.\end{cases}$$

其中$\theta>0$，现随机抽取10台，测得寿命的数据(小时)如下:

1050，1100，1080，1120，1200，1250，1040，1130，1300，1200

试用极大似然估计法求θ的估计值$\hat{\theta}$.

解：似然函数为 $L(\theta)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{x_i}{\theta}}=\frac{1}{\theta^n}e^{-\frac{\sum_{i=1}^{n}x_i}{\theta}}$

上式两边同时求对数得对数似然函数,

$$\ln L=-n\ln\theta-\frac{1}{\theta}\sum_{i=1}^{n}x_i$$

上式两边同时关于θ求导并令其为0,得对数似然方程

$$-\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}x_i=0$$

解方程,得极大似然估计值为

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}$$

把观测数据代入上式,计算得$\hat{\theta}=1147$(小时).

8. 设总体密度函数为

$$f(x;\theta)=\begin{cases}(\theta+1)x^{\theta}, & 0\le x\le 1,\\ 0, & \text{其它.}\end{cases}$$

求参数θ的极大似然估计量和矩估计量

解:似然函数为

$$L(\theta)=\prod_{i=1}^{n}(\theta+1)x_i^{\theta}=(\theta+1)^n\left(\prod_{i=1}^{n}x_i\right)^{\theta}$$

上式两边同时求对数得对数似然函数,

$$\ln L=n\ln(\theta+1)+\theta\sum_{i=1}^{n}\ln x_i$$

上式两边同时关于θ求导并令其为0,得对数似然方程

$$\frac{n}{\theta+1}+\sum_{i=1}^{n}\ln x_i=0$$

解方程，得极大似然估计值为

$$\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln x_i}-1=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\ln x_i^{-1}}-1$$

即θ的极大似然估计量为$\hat{\theta}=\dfrac{1}{\frac{1}{n}\sum_{i=1}^{n}\ln X_i^{-1}}-1$;

$EX=\int_0^1 x(\theta+1)x^{\theta}dx=\dfrac{\theta+1}{\theta+2}$

由$E\hat{X}=\dfrac{\hat{\theta}+1}{\hat{\theta}+2}=\bar{X}$，得$\theta$的矩估计量为$\hat{\theta}=\dfrac{1}{1-\bar{X}}-2$.

9. 随机地从一批零件中抽取 16 个测得其长度(单位: cm)如下:

2.14，2.10，2.13，2.15，2.13，2.12，2.13，2.10

2.15，2.12，2.14，2.10，2.13，2.11，2.14，2.11

设该零件长度分布为正态分布，试求总体均值μ的 0.90 的置信区间.

(1)若已知$\sigma=0.01$；(2)若σ未知.

解:(1)若已知σ，则μ的置信区间为

$$\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

经计算$\bar{x}=2.125$，$\sigma=0.01$，$\alpha=0.1$，经查表得$u_{\alpha/2}=u_{0.05}=1.64$，把数据代入上式得

μ的 0.90 的置信区间为$(2.121,2.129)$.

(2)若σ未知，则μ的置信区间为

$$\bar{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}<\mu<\bar{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}$$

经计算得$\bar{x}=2.125$，$s=0.0171$，又已知$\alpha=0.1$，经查表得$t_{0.05}(15)=1.7531$，把数据代入上式得σ未知时，μ的 0.90 的置信区间为$(2.1175,2.1325)$。

10. 测量某种仪器的工作温度的数据如下

1250℃，1275℃，1265℃，1245℃，1260℃

设温度服从正态分布，在置信水平为 0.95 下，试求温度均值的置信区间.

解：根据题意知，总体方差未知，所以温度均值的置信区间为：

$$\bar{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}<\mu<\bar{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}$$

经计算得 $\bar{x}=1259, s=11.937$，又已知 $\alpha=0.05$，经查表得 $t_{0.025}(4)=2.7765$，把数据代入上式得温度均值的置信区间为(1244.2，1273.8).

11. 设总体 $X\sim N(\mu,\sigma^2)$，已知 $\sum_{i=1}^{15}x_i=8.7$， $\sum_{i=1}^{15}x_i^2=25.05$，试分别求置信水平为 0.95 的 μ 及 σ^2 的置信区间.

解：根据题意，总体均值和方差均未知，所以 μ 的置信区间为：

$$\bar{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}<\mu<\bar{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}$$

经计算得 $\bar{x}=0.58, s=1.195$，又已知 $\alpha=0.05$，经查表得 $t_{0.025}(14)=2.1448$，把数据代入上式得 μ 的置信区间为(－0.08，1.24)；

σ^2 的置信区间为：

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$$

经查表得 $\chi^2_{0.025}(14)=26.119, \chi^2_{0.975}(14)=5.629$，把数据代入上式得 σ^2 的置信区间为：(0.767，3.55).

12. 已知某种木材横纹抗压力的实验值服从正态分布，对 10 个试件作横纹抗压力试验得数据如下(单位：kg/cm^2)：

482，493，457，471，510，446，435，418，394，496

试对该木材横纹抗压力的方差进行区间估计($\alpha=0.05$).

解：由题意知，总体均值未知，要求对总体方差进行区间估计，所以方差的置信区间为：

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$$

经计算得$(n-1)S^2=12439.6$，又已知$\alpha=0.05$，查表得

$\chi^2_{0.025}(9)=19.023,\chi^2_{0.975}(9)=2.700$，把数据代入上式得方差的置信区间为:

(653.924，4607.26).

习题 B

1. 选择题:

(1) 总体X服从参数为λ的指数分布，$X_1,X_2,\cdots,X_n$是样本，$\bar{X}$是样本均值. 若$C\bar{X}^2$为DX的无偏估计，则$C=($　　$)$.

(A) $\frac{n+1}{n}$;　　(B) $\frac{n}{n+1}$;　　(C) $\frac{1}{n}$;　　(D) $\frac{1}{n+1}$.

解: $E\left(C\bar{X}^2\right)=CE\bar{X}^2=C\left(DX+(EX)^2\right)=C\left(\frac{1}{n}\lambda^2+\lambda^2\right)=C\frac{n+1}{n}\lambda^2=\lambda^2$

$\Rightarrow C=\frac{n}{n+1}$，正确答案为(B).

(2) 设总体$X\sim N(\mu,\sigma^2)$ (σ^2已知)，总体均值μ的置信区间的长度L与置信度$1-\alpha$的关系为(　　).

(A) 当$1-\alpha$减少时，L增大;　　(B) 当$1-\alpha$减少时，L缩短;

(C) 当$1-\alpha$减少时，L不变;　　(D) 不能确定.

解: 根据置信区间的性质，当样本容量不变时，置信度越大，置信区间的长度越长. 所以正确答案为(B).

(3) 在正态总体$N(\mu,\sigma^2)$中，μ未知，当用$\left(\bar{X}-z_{0.05}\frac{\sigma}{\sqrt{n}},\bar{X}+z_{0.05}\frac{\sigma}{\sqrt{n}}\right)$作为$\mu$的置信区间时，其置信度为(　　).

(A) 0.1;　　(B) 0.05;　　(C) 0.9;　　(D) 0.95.

解: 由 $\frac{\alpha}{2}=0.05\Rightarrow\alpha=0.1$, 即置信度 $1-\alpha$ 为 0.9, 所以正确答案为 (C).

2. 从正态总体 $N(3.4,6^2)$ 中抽取容量为 n 的样本，如果要求其样本均值位于区间 $(1.4,5.4)$ 内的概率不小于 0.95，问样本容量 n 至少应取多大?

解: 根据题意知，$P\left\{1.4<\bar{X}<5.4\right\}\geq 0.95$, 即

$$P\left\{-\frac{1}{3/\sqrt{n}}<\frac{\bar{X}-3.4}{6/\sqrt{n}}<\frac{1}{3/\sqrt{n}}\right\}\geq 0.95$$

从而推出 $2\Phi\left(\frac{\sqrt{n}}{3}\right)-1\geq 0.95\Rightarrow\Phi\left(\frac{\sqrt{n}}{3}\right)\geq 0.975$, 经查表得 $n>34.6$, 即 n 至少应取 35.

3. 设总体 X 的密度函数为 $f(x;\theta)$，$X_1,X_2,\cdots,X_n$ 为其样本，求参数 θ 的极大似然估计:

(1) $f(x;\theta)=\frac{1}{2}\mathrm{e}^{-|x-\theta|}$，$|x|>\theta$，$-\infty<\theta<\infty$，$-\infty<x<\infty$.

(2) $f(x;\theta)=\begin{cases}\theta x^{\theta-1}, & 0<x<1, 0<\theta<\infty,\\ 0, & \text{其它}.\end{cases}$

(3) $f(x;\theta)=\begin{cases}\frac{1}{\theta}, & 0<x\leq\theta, 0<\theta<\infty,\\ 0, & \text{其它}.\end{cases}$.

解: (1) 似然函数为 $L(\theta)=\prod_{i=1}^{n}\frac{1}{2}e^{-|x_i-\theta|}=\frac{1}{2^n}e^{-\sum_{i=1}^{n}|x_i-\theta|}$, 要使得似然函数取极大值, 即要使得 $\sum_{i=1}^{n}|x_i-\theta|$ 取极小值, 又根据题意 $|x|>\theta$，$-\infty<\theta<\infty$，$-\infty<x<\infty$, 所以, 若样本的顺序统计量记为 $X_1^*,X_2^*,\cdots,X_n^*$, 不难看出:

当 $n=2k+1$ 时，$\hat{\theta}=X_{k+1}^*$ 使得 $\sum_{i=1}^{n}|x_i-\theta|$ 极小;

当 $n=2k$ 时，$X_k^*<\hat{\theta}<X_{k+1}^*$ 使得 $\sum_{i=1}^{n}|x_i-\theta|$ 极小.

(2) 似然函数为 $L(\theta)=\prod_{i=1}^{n}\theta x_i^{\theta-1}=\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}$,

等式两边同时求对数, 得对数似然函数为 $\ln L(\theta)=n\ln\theta+(\theta-1)\sum_{i=1}^{n}\ln x_i$,

等式两边同时关于θ求导得，对数似然方程$\frac{n}{\theta}+\sum_{i=1}^{n}\ln x_i=0$

解方程得θ的极大似然函数为$\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln X_i}$；

(3) 似然函数为$L(\theta)=\prod_{i=1}^{n}\theta=\theta^n, 0<x_i<\theta, i=1,2,\cdots,n$

等式两边同时求对数，得对数似然函数为$\ln L(\theta)=n\ln\theta, 0<x_i<\theta, i=1,2,\cdots,n$，

若样本的顺序统计量记为$X_1^*, X_2^*,\cdots, X_n^*$，不难看出：

当$\hat{\theta}=X_n^*=\max\limits_{1\le i\le n}X_i$时，对数似然函数取极大值，即$\theta$的极大似然函数为$\hat{\theta}=\max\limits_{1\le i\le n}X_i$。

4. 为了比较 *A*，*B* 两种灯泡的寿命，从 *A* 型号中随机地抽取 80 只，测得平均寿命$\overline{x}=2000$（小时），样本标准差$s_1=80$（小时），从 *B* 型号中随机抽取 100 只，测得平均寿命$\overline{y}=1900$（小时），样本标准差$s_2=100$（小时），假定两种型号的灯泡寿命分别服从正态分布$N(\mu_1,\sigma_1^2)$和$N(\mu_2,\sigma_2^2)$且相互独立，试求：

(1) 置信度为 0.99 的$\mu_1-\mu_2$的置信区间；

(2) $\frac{\sigma_1^2}{\sigma_2^2}$的置信度为 0.90 的置信区间.

解：(1) 由于两总体的样本容量都比较大，所以$\mu_1-\mu_2$的置信区间为：

$$\left((\overline{X}-\overline{Y})-u_{\alpha/2}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}},(\overline{X}-\overline{Y})+u_{\alpha/2}\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}\right)$$

经查表得$u_{0.005}=2.57$，把数据代入上式得$\mu_1-\mu_2$的置信度为 0.99 的置信区间为 (65.4，134.6)；

(2) $\frac{\sigma_1^2}{\sigma_2^2}$的一个置信度为$1-\alpha$的置信区间为

$$\left(\frac{S_1^2/S_2^2}{F_{\alpha/2}(n-1,m-1)},\frac{S_1^2/S_2^2}{F_{1-\alpha/2}(n-1,m-1)}\right)$$

经查表得 $F_{0.05}(79,99)\approx 1.4298$，$F_{0.95}(79,99)=\dfrac{1}{F_{0.05}(99,79)}\approx 0.6803$，将数据代入上式得 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信度为 0.90 的置信区间为(0.4476，0.9408).

5. 设某种清漆的 9 个样本，其干燥时间(单位：小时)分别为

6.0，5.7，5.8，6.5，7.0，6.3，5.6，6.1，5.0

设干燥时间总体服从正态分布 $N(\mu,\sigma^2)$，求 μ 的置信水平为 0.95 的单侧置信上限，(1)若由以往经验知 $\sigma=0.6$ (h)；(2)若 σ 为未知.

解：(1)若已知 $\sigma=0.6$，则 μ 的置信水平为 0.95 的单侧置信上限为

$$\bar{\mu}=\bar{X}+\frac{\sigma}{\sqrt{n}}u_\alpha=6.0+0.6/3*1.64=6.328$$

(2)若 σ 为未知，则 μ 的置信水平为 0.95 的单侧置信上限为

$$\bar{\mu}=\bar{X}+\frac{S}{\sqrt{n}}t_\alpha(n-1)=6.0+0.5745/3*1.8595=6.356$$

6. 设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim N(\mu,\sigma^2)$ 的样本，μ 已知. 问 σ^2 的两个无偏估计量 $S_1^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$ 和 $S_2^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 哪个更有效?

解：$\dfrac{X_i-\mu}{\sigma}\sim N(0,1)$，$i=1,2,\cdots,n$，则

$$\frac{nS_1^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$$

所以，$D\left(\dfrac{nS_1^2}{\sigma^2}\right)=\dfrac{n^2}{\sigma^4}D\left(S_1^2\right)=2n\Rightarrow D\left(S_1^2\right)=\dfrac{2n\sigma^4}{n^2}=\dfrac{2\sigma^4}{n}$

又已知 $\dfrac{(n-1)S_2^2}{\sigma^2}\sim\chi^2(n-1)$，所以

$$D\left(\frac{(n-1)S_2^2}{\sigma^2}\right)=\frac{(n-1)^2}{\sigma^4}D\left(S_2^2\right)=2(n-1)\Rightarrow D\left(S_2^2\right)=\frac{2(n-1)\sigma^4}{(n-1)^2}=\frac{2\sigma^4}{n-1}$$

不难判断，$D\left(S_1^2\right)<D\left(S_2^2\right)$，即 S_1^2 比 S_2^2 更有效.

7．设某人上班的等车时间 X 服从$[0,\theta]$上的均匀分布，θ为未知参数，现有样本 $x_1=4.2$，$x_2=3.5$，$x_3=1.7$，$x_4=1.2$，$x_5=2.4$，求θ的置信度为 0.95 的置信区间.

解：X 服从均匀分布，所以$EX=\dfrac{\theta}{2}$，又因为总体方差未知，则$\dfrac{\bar{X}-\frac{\theta}{2}}{S/\sqrt{n}}\sim N(0,1)$，

所以$\dfrac{\theta}{2}$的置信区间为：

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

经计算得$\bar{x}=2.6,s=1.2430$，又经查表得$t_{0.025}(4)=2.7765$，把数据代入上式计算得θ的置信度为 0.95 的置信区间为$(2.11,4.14)$.

第 8 章　假设检验

习题 A

1．据经验知某零件重量 $X \sim N(\mu,\sigma^2)$，$\mu = 15$，$\sigma^2 = 0.05^2$，技术革新后，抽检 6 个零件，测得其重量为(单位：g)

14.7，15.1，14.8，15.0，15.2，14.6

已知方差不变，试问平均重量是否仍为 15g($\alpha = 0.05$)？

解：由题意知，需检验假设：

$$H_0: \mu = \mu_0 = 15, H_1: \mu \neq 15$$

总体方差 $\sigma^2 = 0.05^2$ 已知，所以选择统计量 $U = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$，得拒绝域为

$$|u| = \left|\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}\right| \geq u_{\alpha/2}$$

将 $n = 6, u_{\alpha/2} = 1.96, \sigma = 0.05, \overline{x} = 14.9$，代入上式计算得 $|u| = 4.899 > 1.96$，所以拒绝原假设，即可认为零件的平均重量不再是 15g.

2．设某厂生产的一种钢索，其断裂强度 X (kg/cm^2) 服从正态分布 $N(\mu, 40^2)$，从中选取一个容量为 9 的样本，得 $\overline{x} = 780$ (kg/cm^2)．在显著性水平 $\alpha = 0.05$ 下可否据此认为这批钢索的断裂强度为 800kg/cm^2?

解: 由题意知，需检验假设:

$$H_0: \mu = \mu_0 = 800, H_1: \mu \neq 800$$

总体方差 $\sigma^2 = 40^2$ 已知，所以选择统计量 $U = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$，得拒绝域为

$$|u| = \left|\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}\right| \geq u_{\alpha/2}$$

将 $n = 9, u_{\alpha/2} = 1.96, \sigma = 40, \overline{x} = 780$，代入上式计算得 $|u| = 1.5 < 1.96$，所以接受原假设，即可认为这批钢索的断裂强度为 800kg/cm^2.

3. 某地早稻收割根据长势估计平均亩产为 310kg，收割时，随机抽取了 10 块，测出每块的实际亩产量为 $x_1,x_2,\cdots,x_n$. 计算得 $\bar{x}=320\text{kg}$. 如果已知早稻亩产量 X 服从正态分布 $N(\mu,12^2)$，试问所估产量是否正确？（$\alpha=0.05$）

解：由题意知，需检验假设：

$$H_0:\mu=\mu_0=310, H_1:\mu\neq 310$$

总体方差 $\sigma^2=12^2$ 已知，所以选择统计量 $U=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$，得拒绝域为

$$|u|=\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right|\geq u_{\alpha/2}$$

将 $n=10, u_{\alpha/2}=1.96, \sigma=12, \bar{x}=320$，代入上式计算得 $|u|=2.635>1.96$，所以拒绝原假设，即可认为所估产量是不正确的.

4. 如果一个矩形的宽度 w 与长度 l 的比 $\dfrac{w}{l}=\dfrac{1}{2}(\sqrt{5}-1)\approx 0.618$，这样的矩形称为黄金矩形. 这种尺寸的矩形使人们看上去有良好的感觉. 现代的建筑构件（如窗架）、工艺品（如图片镜框），甚至司机的执照、商业的信用卡等常常都是采用黄金矩形. 下面是某工艺品随机抽取的 20 个矩形的宽度与长度的比值数据：

0.693	0.749	0.654	0.670	0.662	0.672	0.615	0.606	0.690	0.628
0.668	0.611	0.606	0.609	0.601	0.553	0.570	0.844	0.576	0.933

设这一工厂生产的矩形的宽度与长度的比值总体服从正态分布，其均值为 μ，方差为 σ^2，两者均未知. 试检验假设（取 $\alpha=0.05$）

$$H_0:\mu=0.618;\quad H_1:\mu\neq 0.618$$

解：由题意知，需检验假设：

$$H_0:\mu=0.618;\quad H_1:\mu\neq 0.618$$

总体方差未知，所以选择统计量 $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$，得拒绝域为

$$|t|=\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|\geq t_{\alpha/2}(n-1)$$

将 $n=20, t_{0.025}(19)=2.0930, s=0.092511, \bar{x}=0.6605$，代入上式得 $|t|=2.0545<2.0930$，所以接受原假设.

5. 设某次考试的学生成绩服从正态分布，从中随机地抽取 36 位考生的成绩，算得平均成绩为 66.5 分，标准差为 15 分，问在显著性水平 α=0.05 下，是否可以认为这次考试考生的平均成绩为 70 分？并给出检验过程.

解：由题意知，需检验假设：

$$H_0: \mu=70; \quad H_1: \mu \neq 70$$

总体方差未知，所以选择统计量 $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$，得拒绝域为

$$|t|=\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \geq t_{\alpha/2}(n-1)$$

将 $n=36, t_{0.025}(35)=2.0301, s=15, \bar{x}=66.5$，代入上式得 $|t|=1.3804<2.0301$，所以接受原假设，即可认为这次考试考生的平均成绩为 70 分.

6. 8 名学生独立地测定同一物质的比重，分别测得其值（单位：g/cm^3）为

11.49，11.51，11.52，11.53，11.47，11.46，11.55，11.50

假定测定值服从正态分布，试根据这些数据检验该物质的实际比重是否为 11.53（$\alpha=0.05$）.

解：由题意知，需检验假设：

$$H_0: \mu=11.53; \quad H_1: \mu \neq 11.53$$

总体方差未知，所以选择统计量 $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$，得拒绝域为

$$|t|=\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \geq t_{\alpha/2}(n-1)$$

将 $n=8, t_{0.025}(7)=2.3646, s=0.0302, \bar{x}=11.504$，代入上式得 $|t|=2.4579>2.3646$，所以拒绝原假设，即可认为该物质的实际比重不是 11.53.

7. 已知健康人的红血球直径服从均值为 7.2 μm 的正态分布，今在某患者血液中随机测得 9 个红血球的直径如下：

7.8，9.0，7.1，7.6，8.5，7.7，7.3，8.1，8.0

问该患者红血球平均直径与健康人的差异是否显著不同（$\alpha=0.05$）?

解：由题意知，需检验假设:

$$H_0:\mu=7.2;\quad H_1:\mu\neq 7.2$$

总体方差未知，所以选择统计量 $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$，得拒绝域为

$$|t|=\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|\geq t_{\alpha/2}(n-1)$$

将 $n=9, t_{0.025}(8)=2.3060, s=0.5874, \bar{x}=7.9$，代入上式得 $|t|=3.5753>2.3060$，所以拒绝原假设，即可认为患者红血球平均直径与健康人的差异是显著不同的.

8. 某纺织厂在正常条件下，平均每台织布机每小时经纱断头根数为 9.73 根. 该厂进行工艺改革，减少经纱上浆率. 在 100 台织布机上进行试验，结果平均每台每小时断头根数为 9.94，标准差为 1.20. 已知经纱断头根数服从正态分布，试在水平 $\alpha=0.05$ 下检验新的上浆率能否推广使用?（一般情况下，上浆率降低，断头根数应增加）

解：由题意知，需检验假设:

$$H_0:\mu\leq 9.73;\quad H_1:\mu>9.73$$

总体方差未知，所以选择统计量 $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$，得拒绝域为

$$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\geq t_{\alpha}(n-1)$$

将 $n=100, t_{0.05}(99)\approx u_{0.05}=1.64, s=1.20, \bar{x}=9.94$，代入上式得 $t=0.525<1.64$，所以接受原假设，即可认为新的上浆率不能推广使用.

9. 某厂生产的某种型号的电池，其寿命（单位：小时）长期以来服从方差 $\sigma^2=5000$ 的正态分布，现有一批这种电池，从它的生产情况来看，寿命的波动性有所改变. 现随机取 26 只电池，测出其寿命的样本方差 $s^2=9200$. 问根据这一数据能否推断这批电池的寿命的波动性较以往的有显著的变化（$\alpha=0.05$）?

解: 由题意知, 需检验假设

$$H_0:\sigma^2=\sigma_0^2=5000;H_1:\sigma^2\neq 5000$$

总体均值未知，所以选择统计量 $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$，当 H_0 为真时，$\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$, 从而拒绝域为

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\leq\chi^2_{1-\alpha/2}(n-1) \text{ 或 } \chi^2=\frac{(n-1)S^2}{\sigma_0^2}\geq\chi^2_{\alpha/2}(n-1)$$

将 $n=26,\alpha=0.05,s^2=9200,\sigma_0^2=5000$, 代入检验统计量得 $\chi^2\approx 46$, 查表得 $\chi^2_{0.025}(25)=40.646,\chi^2_{0.975}(25)=13.120$, 即有 $\chi^2\approx 46>\chi^2_{0.025}(25)$, 所以拒绝原假设, 即可认为这批电池的寿命的波动性较以往有显著的变化.

10. 无线电厂生产的某种高频管，其中一项指标服从正态分布 $N(\mu,\sigma^2)$，今从一批产品中抽取 8 个高频管，测得指标数据为

68，43，70，65，55，56，60，72

(1) 已知总体期望 $\mu=60$ 时，检验假设 $H_0:\sigma^2=8^2$（$\alpha=0.05$）;

(2) 总体期望 μ 未知时，检验假设 $H_0:\sigma^2=8^2$（$\alpha=0.05$）.

解: (1) 由题意知, 需检验假设

$$H_0:\sigma^2=\sigma_0^2=8^2;H_1:\sigma^2\neq 8^2$$

总体期望已知，故选择统计量 $\chi^2=\sum_{i=1}^{n}\dfrac{(X_i-\mu)^2}{\sigma_0^2}$，当 H_0 为真时，$\sum_{i=1}^{n}\dfrac{(X_i-\mu)^2}{\sigma_0^2}\sim\chi^2(n)$, 从而拒绝域为

$$\chi^2=\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma_0^2}\leq\chi^2_{1-\alpha/2}(n-1) \text{ 或 } \chi^2=\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma_0^2}\geq\chi^2_{\alpha/2}(n-1)$$

将 $n=8,\alpha=0.05,\sum_{i=1}^{n}(X_i-\mu)^2=654,\sigma_0^2=64$, 代入检验统计量计算得 $\chi^2\approx 10.2188$, 查表得 $\chi^2_{0.025}(8)=17.535,\chi^2_{0.975}(8)=2.180$, 即有 $\chi^2_{0.975}(8)<\chi^2<\chi^2_{0.025}(8)$, 所以接受

原假设，即认为总体方差为8^2；

(2) 由题意知，需检验假设

$$H_0:\sigma^2=\sigma_0^2=8^2;H_1:\sigma^2\neq 8^2$$

总体均值未知，所以选择统计量 $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$，当 H_0 为真时，$\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$，从而拒绝域为

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\le\chi^2_{1-\alpha/2}(n-1) \text{ 或 } \chi^2=\frac{(n-1)S^2}{\sigma_0^2}\ge\chi^2_{\alpha/2}(n-1)$$

将 $n=8,\alpha=0.05,s^2=93.2679,\sigma_0^2=64$，代入检验统计量得 $\chi^2\approx 10.2012$，查表得 $\chi^2_{0.025}(7)=16.013,\chi^2_{0.975}(7)=1.690$，即有 $\chi^2_{0.975}(7)<\chi^2<\chi^2_{0.025}(7)$，所以接受原假设，即认为总体方差为$8^2$.

11. 正常成年人的脉搏平均值为 72 次/分钟，现有某一医生为研究慢性四乙基铅中毒者的脉搏平均值与正常成年人的脉搏平均值的关系，他随机地抽查了 10 例患者的脉搏如下:

54，67，78，68，70，66，70，67，65，69

设脉搏服从正态分布，问:

(1) 慢性四乙基铅中毒者的脉搏平均值与正常成年人的脉搏平均值有无显著差异?

(2) 若正常成年人脉搏的方差为 $\sigma_0^2=36$，那么他们两者之间的方差有无显著差异?（$\alpha=0.05$）

解: (1) 由题意知，需检验假设:

$$H_0:\mu=72;\quad H_1:\mu\neq 72$$

总体方差未知，所以选择统计量 $T=\dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$，得拒绝域为

$$|t|=\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|\ge t_{\alpha/2}(n-1)$$

将 $n=10, t_{0.025}(9)=2.2622, s=5.9292, \bar{x}=67.4$，代入上式得 $|t|=2.4534>2.2622$，所以拒绝原假设，即可认为慢性四乙基铅中毒者的脉搏平均值与正常成年人的脉搏平均值有显著差异；

(2) 由题意知，需检验假设

$$H_0:\sigma^2=\sigma_0^2=36; H_1:\sigma^2\neq 36$$

总体均值未知，所以选择统计量 $\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$，当 H_0 为真时，$\frac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$，从而拒绝域为

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\leq\chi^2_{1-\alpha/2}(n-1) \text{ 或 } \chi^2=\frac{(n-1)S^2}{\sigma_0^2}\geq\chi^2_{\alpha/2}(n-1)$$

将 $n=10, \alpha=0.05, (n-1)s^2=316.4, \sigma_0^2=36$，代入检验统计量得 $\chi^2\approx 8.7889$，查表得 $\chi^2_{0.025}(9)=19.023, \chi^2_{0.975}(9)=2.700$，即有 $\chi^2_{0.975}(9)<\chi^2<\chi^2_{0.025}(9)$，所以接受原假设，即可认为两者的脉搏方差无显著差异。

习题 B

1. 随机地挑选 20 位失眠者，分别使用甲、乙两种安眠药，记录下他们的睡眠的延长时间(单位：小时)，得如下数据：

服用甲药	1.9	0.8	1.1	-0.1	0.1	4.4	5.6	1.6	4.6	3.4
服用乙药	0.7	-1.6	-0.2	-0.1	3.4	3.7	0.8	-1.2	2.0	0

μ_1,μ_2 分别表示服用甲药和乙药的失眠者的平均睡眠延长时间。试问能否认为甲药的疗效显著高于乙药，即需检验假设

$$H_0:\mu_1\leq\mu_2;\quad H_1:\mu_1>\mu_2$$

(假设总体服从正态分布，取显著性水平 $\alpha=0.05$)。

解: 由题意知，需检验假设

$$H_0:\mu_1\le\mu_2;\quad H_1:\mu_1>\mu_2$$

由于假设两总体方差相等但未知, 所以可以选择检验统计量为

$$T=\frac{(\bar{X}-\bar{Y})}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

当 H_0 为真时，$T=\dfrac{(\bar{X}-\bar{Y})}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t(n+m-2)$，从而可以确定拒绝域为

$$t=\frac{\bar{x}-\bar{y}}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\ge t_\alpha(n+m-2)$$

这里，$n=m=10,t_{0.05}(18)=1.7341,\bar{x}=2.34,\bar{y}=0.75,s_\omega\approx1.9080$，代入检验统计量得 $t=1.8633>t_{0.05}(18)$，即拒绝原假设，可认为甲药的疗效显著高于乙药的疗效.

2. 在 20 世纪 70 年代后期人们发现，在酿造啤酒时，麦芽干燥过程中会形成致癌物质亚硝基二甲胺（*NDMA*）. 到了 80 年代初开发了一种新的麦芽干燥过程. 下面给出分别在新老两种过程中形成 *NDMA* 含量（以 10 亿份中的份数计）:

新过程	6	4	5	5	6	5	5	6	4	6	7	4
老过程	2	1	2	2	1	0	3	2	1	0	1	3

设两样本分别来自正态总体，且两总体的方差相等，两样本相互独立. 分别以 μ_1，μ_2 表示老、新过程的总体均值，试检验假设（取 $\alpha=0.05$）

$$H_0:\mu_1-\mu_2\le2;\quad H_1:\mu_1-\mu_2>2.$$

解: 由题意知, 检验假设

$$H_0:\mu_1-\mu_2\le2;\quad H_1:\mu_1-\mu_2>2.$$

由于假设两总体方差相等但未知，所以可以选择检验统计量为

$$T=\frac{(\bar{X}-\bar{Y})-\delta}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}.$$

当 H_0 为真时，$T=\dfrac{(\bar{X}-\bar{Y})-\delta}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t(n+m-2)$，从而可以确定拒绝域为

$$t=\frac{\bar{x}-\bar{y}-\delta}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\geq t_\alpha(n+m-2)$$

这里， $n=12, m=12, t_\alpha(n+m-2)=t_{0.05}(22)=1.7171, \delta=2$ ， 由样本算得 $\bar{x}=5.25$， $\bar{y}=1.5$， $s_\omega\approx 0.9828$， 于是

$t=4.1759>t_{0.05}(22)$，所以拒绝原假设.

3. 两种小麦品种从播种到抽穗所需的天数如下:

x	101	100	99	99	98	100	98	99	99	99
y	100	98	100	99	98	99	98	98	99	100

设两样本依次来自正态总体 $N(\mu_1,\sigma_1^2)$， $N(\mu_2,\sigma_2^2)$， μ_i， σ_i^2（$i=1,2$）均未知，两样本相互独立.

(1) 试检验假设 $H_0:\sigma_1^2=\sigma_2^2$， $H_1:\sigma_1^2\neq\sigma_2^2$（取 $\alpha=0.05$）;

(2) 若能接受 H_0，接着检验假设 $H_0:\mu_1=\mu_2$， $H_1:\mu_1\neq\mu_2$（取 $\alpha=0.05$）.

解: (1) 由题意知，检验假设:

$$H_0:\sigma_1^2=\sigma_2^2,\quad H_1:\sigma_1^2\neq\sigma_2^2$$

选取检验统计量

$$F=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}.$$

当 H_0 为真时，$\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n-1,m-1)$，从而可以确定拒绝域为

$$F=\frac{S_1^2}{S_2^2}\geq F_{\alpha/2}(n-1,m-1) \text{ 或 } F=\frac{S_1^2}{S_2^2}\leq F_{1-\alpha/2}(n-1,m-1)$$

其中

$n=m=10,\alpha=0.05,F_{0.025}(9,9)=4.03,F_{0.975}(9,9)=\dfrac{1}{F_{0.025}(9,9)}=\dfrac{1}{4.03}\approx 0.2481$ ，

$F=\dfrac{S_1^2}{S_2^2}\approx 1.0145$，不难看出 $F_{0.975}(9,9)<F<F_{0.025}(9,9)$，所以接受原假设.

(2) 由题意知，检验假设

$$H_0:\mu_1=\mu_2，\quad H_1:\mu_1\neq\mu_2$$

由于假设两总体方差相等但未知，所以可以选择检验统计量为

$$|T|=\frac{|(\bar{X}-\bar{Y})-\delta|}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}.$$

当 H_0 为真时，$T=\dfrac{(\bar{X}-\bar{Y})-\delta}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t(n+m-2)$，从而可以确定拒绝域为

$$|t|=\frac{|\bar{x}-\bar{y}-\delta|}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\geq t_{\alpha/2}(n+m-2)$$

其中 $n=m=10,\alpha=0.05,t_{0.025}(18)=1.7341,\delta=0$，经计算得

$\bar{x}=99.2,\bar{y}=98.9,s_\omega\approx 0.8975$，于是 $|t|=0.7474<t_{0.025}(18)$，所以接受原假设.

4. 甲、乙两家农业银行分别对 21 个储户和 16 个储户的年存款余额进行抽样检查，测得其平均年存款余额分别为 $\bar{x}=2600$ 元和 $\bar{y}=2700$，样本方差相应为 $s_1^2=81^2$ 和 $s_2^2=105^2$. 假设年存款余额服从正态分布，试在显著性水平 $\alpha=0.05$ 下检验假设

$$H_0:\sigma_1^2\leq\sigma_2^2，\quad H_0:\sigma_1^2>\sigma_2^2.$$

解: 由题意知，检验假设

$$H_0:\sigma_1^2\leq\sigma_2^2，\quad H_0:\sigma_1^2>\sigma_2^2.$$

选取检验统计量

$$F=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}.$$

当 H_0 为真时，$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n-1,m-1)$，从而可以确定拒绝域为

$$F=\frac{S_1^2}{S_2^2}\geq F_\alpha(n-1,m-1)$$

其中 $n=21,m=16,\alpha=0.05,F_{0.05}(20,15)=2.33,$ 经计算得

$F=\frac{S_1^2}{S_2^2}\approx 0.5951<F_{0.05}(20,15)$，接受原假设，即认为银行甲的年存款余额方差小于银行乙.

5. 根据某市公路交通部门某年中前 6 个月交通事故记录，统计得星期一至星期日发生交通事故的次数如下:

星期	1	2	3	4	5	6	7(日)
次数	36	23	29	31	34	60	25

问交通事故是否均匀地发生在一周 7 天中(取 $\alpha=0.05$)?

解: 若事故均匀发生在 7 天镍，则应有 $p_i=\frac{1}{7}$，$i=1,2,\cdots,7$，以 A_i 表示事件“事故发生在周 i”，则需要检验假设

$$H_0:P(A_i)=\frac{1}{7}\ (i=1,2,\cdots,7).$$

选择检验统计量 $\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}$，由定理 1 知

$$\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\sim\chi^2(r-1).$$

当显著性水平为 $\alpha=0.05$ 时，检验的拒绝域为

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}\geq\chi_\alpha^2(r-1).$$

又 $n=\sum_{i=1}^{r}n_i=238,np_i=238\times\frac{1}{7}=34$，所以

$$\chi^2=\sum_{i=1}^{7}\frac{(n_i-np_i)^2}{np_i}=\frac{(36-34)^2}{34}+\frac{(23-34)^2}{34}+\frac{(29-34)^2}{34}+$$

$$\frac{(31-34)^2}{34}+\frac{(34-34)^2}{34}+\frac{(60-34)^2}{34}+\frac{(25-34)^2}{34}\approx 26.94$$

经查表得 $\chi^2_{0.05}(6)=12.592$，显然 $26.94>12.592$，表明样本观察值落在拒绝域内.

因此在 $\alpha=0.05$ 下拒绝 H_0，即认为交通事故并不是均匀发生在一周 7 天中.

6. 测量 100 根人造纤维的长度(单位：mm)，所得数据如下表所示：

长度	5.5～6.0	6.0～6.5	6.5～7.0	7.0～7.5	7.5～8.0	8.0～8.5
频数	2	7	6	17	17	16
长度	8.5～9.0	9.0～9.5	9.5～10.0	10.0～10.5	10.5～11	
频数	14	10	7	3	1	

问能否认为人造纤维长度服从正态分布(取 $\alpha=0.05$)？

解：记 X 为人造纤维的长度，由题意需检验

$$H_0:\quad X\sim N(\mu,\sigma^2).$$

由于 H_0 中含有两个未知参数，因此需先进行参数估计. μ 和 σ^2 的极大似然估计值分别为（x_i 表示组中值，f_i 表示频数）

$$\hat{\mu}=\bar{x}=\frac{\sum_{i=1}^{n}x_i f_i}{\sum_{i=1}^{n}f_i}$$

$$=[2\times5.75+7\times6.25+6\times6.75+17\times7.25+17\times7.75+16\times8.25+$$
$$14\times8.75+10\times9.25+7\times9.75+3\times10.25+1\times10.75]/100$$
$$=8.075.$$

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i}{\sum_{i=1}^{n} f_i}$$

$$=[2\times(5.75-8.075)^2+7\times(6.25-8.075)^2+6\times(6.75-8.075)^2+$$
$$17\times(7.25-8.075)^2+17\times(7.75-8.075)^2+16\times(8.25-8.075)^2+$$
$$14\times(8.75-8.075)^2+10\times(9.25-8.075)^2+7\times(9.75-8.075)^2+$$
$$3\times(10.25-8.075)^2+1\times(10.75-8.075)^2]/100$$
$$=1.196875.$$

由于 $\hat{p}_1=\hat{P}\{X\le 6.5\}=\Phi\left(\frac{6.5-\hat{\mu}}{\hat{\sigma}}\right)$，且

$$\hat{p}_i=\hat{P}\{\alpha_{i-1}<X\le\alpha_i\}=\Phi\left(\frac{\alpha_i-\hat{\mu}}{\hat{\sigma}}\right)-\Phi\left(\frac{\alpha_{i-1}-\hat{\mu}}{\hat{\sigma}}\right),\quad i=2,3,\cdots,7$$

$$\hat{p}_8=\hat{P}\{9.5<X<\infty\}=1-\Phi\left(\frac{9.5-\hat{\mu}}{\hat{\sigma}}\right).$$

结果如下表:

A_i	n_i	$\hat{p}_i$	$n\hat{p}_i$	$n_i-n\hat{p}_i$	$\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}$
$X\le 6.5$	9	0.075	7.5	1.5	0.3
$6.5<X\le 7$	6	0.088	8.8	-2.8	0.8909
$7<X\le 7.5$	17	0.137	13.7	3.3	0.79489
$7.5<X\le 8$	17	0.174	17.4	-0.4	0.0092
$8<X\le 8.5$	16	0.1758	17.58	-1.58	0.142
$8.5<X\le 9$	14	0.1511	15.11	-1.11	0.0815
$9<X\le 9.5$	10	0.1023	10.23	-.23	0.0517
$9.5<X$	11	0.0968	9.68	1.32	0.18
合计					2.45019

从而 $r=8$， $\chi^2_\alpha(r-k-1)=\chi^2_{0.05}(8-2-1)=11.07>2.45019$. 因此接受原假设

H_0，认为人造纤维长度服从正态分布.

7. 下面列出了 84 个伊特拉斯坎(Etruscan)人男子的头颅的最大宽度(单位: mm):

141	148	132	138	154	142	150	146	155	158
150	140	147	148	144	150	149	145	149	158
143	141	144	144	126	140	144	142	141	140
145	135	147	146	141	136	140	146	142	137
148	154	137	139	143	140	131	143	141	149
148	135	148	152	143	144	141	143	147	146
150	132	142	142	143	153	149	146	149	138
142	149	142	137	134	144	146	147	140	142
140	137	152	145						

试检验它们是否来自正态总体（取$\alpha = 0.1$）.

解: 记 X 为伊特拉斯坎(Etruscan)人男子的头颅的最大宽度, 依题意检验

$$H_0: \quad X \sim N(\mu, \sigma^2)$$

由于H_0中含有两个未知参数，因此需先进行参数估计. μ和σ^2的极大似然估计值分别为

$$\hat{\mu} = \overline{x} = 143.8, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = (6.0)^2$$

我们将在H_0下 X 可能取值的区间$(-\infty, \infty)$分为 7 个小区间, 并取事件A_i如下表所示. 若H_0为真 X 的概率密度的估计为

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi} \times 6} e^{-\frac{(x-143.8)^2}{2\times 6^2}}, \quad -\infty < x < \infty$$

按上式并查标准正态分布的分布函数表即可得概率$P(A_i)$的估计. 例如

$$\hat{p}_i = \hat{P}(A_i) = \hat{P}\{129.5 < X \le 134.5\} =$$

$$\Phi\left(\frac{134.5 - 143.8}{6}\right) - \Phi\left(\frac{129.5 - 143.8}{6}\right) = \Phi(-1.55) - \Phi(-2.38) = 0.0519$$

将计算结果列表如下:（χ^2 检验计算表）

A_i	f_i	$\hat{p}_i$	$n\hat{p}_i$	$f_i^2/n\hat{p}_i$
$A_1: x \le 129.5$	$\left.\begin{matrix}1\\4\end{matrix}\right\}5$	$\left.\begin{matrix}0.0087\\0.0519\end{matrix}\right\}$	$\left.\begin{matrix}0.73\\4.36\end{matrix}\right\}5.09$	4.91
$A_2: 129.5 < x \le 134.5$				
$A_3: 134.5 < x \le 139.5$	10	0.1752	14.72	6.79
$A_4: 139.5 < x \le 144.5$	33	0.3120	26.21	41.55
$A_5: 144.5 < x \le 149.5$	24	0.2811	23.61	24.40
$A_6: 149.5 < x \le 154.5$	$\left.\begin{matrix}9\\3\end{matrix}\right\}12$	$\left.\begin{matrix}0.1336\\0.0375\end{matrix}\right\}$	$\left.\begin{matrix}11.22\\3.15\end{matrix}\right\}14.37$	10.02
$A_7: 154.5 < x < \infty$				
合计				87.67

现在 $\chi^2 = 87.67 - 84 = 3.67$，因为

$\chi^2_{0.1}(k-r-1) = \chi^2_{0.1}(5-2-1) = \chi^2_{0.1}(2) = 4.605 > 3.67$，故在水平 $\alpha = 0.1$ 下接受原假设，即认为数据来自正态分布总体.

第9章　方差分析与回归分析

习题 A

1. 考察温度对某一化工产品得率的影响，选了五种不同的温度，在同一温度下做了三次实验，测得其得率如下，试分析温度对得率有无显著影响.

温度	60	65	70	75	80
得率	90	91	96	84	84
	92	93	96	83	89
	88	92	93	83	82

解 把原始数据均减去90后可列出如下计算表和方差分析表，r 表示因子水平数，t 为重复实验次数.

$$r=5, t=3, n=rt=15$$

计算表

温度	60	65	70	75	80	
y_{ij}	0	1	6	−6	−6	
	2	3	6	−7	−1	
	−2	2	3	−7	−8	
$y_{i\cdot}$	0	6	15	−20	−15	$\sum_i y_i=-14$

$$\sum_i\sum_j y_{ij}^2=338, \sum_i y_i^2=886, \frac{\left(\sum_i\sum_j y_{ij}\right)^2}{n}\approx 13.067$$

$$S_A=\frac{1}{3}\times 886-13.067\approx 282.266$$

$$S_T=338-13.067=324.933$$

$$S_e=S_T-S_A=42.667$$

方差分析表

来源	平方和	自由度	均方和	F 比
温度	282.266	4	70.6	16.4
e	42.667	10	4.3	
总和	324.933	17	$F_{0.99}(4,10)=6$	

所以，在$\alpha=0.01$上水平上认为温度对得率有显著影响.

2. 下面记录了三位操作工分别在四台不同机器上操作三天的日产量:

机 器	操作工								
	甲			乙			丙		
A_1	15	15	17	17	19	16	16	18	21
A_2	17	17	17	15	15	15	19	22	22
A_3	15	17	16	18	17	16	18	18	18
A_4	18	20	22	15	16	17	17	17	17

试在显著性水平$\alpha=0.05$下检验:

(1) 操作工之间有无显著性差异?

(2) 机器之间的差异是否显著?

(3) 操作工与机器的交互作用是否显著?

解 用r表示机器的水平数，s表示操作工的水平数，t表示重复实验次数，列出计算表和方差分析表:

$$r=4,s=3,t=3,n=rst=36$$

y_{ij}	甲	乙	丙	$y_{i.}$
A_1	47	52	55	154
A_2	51	45	63	159
	48	51	54	153
A_3	60	48	51	159

A_4				
$y_{.j}$	206	196	223	625

$n=36$，$\sum_i\sum_j\sum_k y_{ijk}^2=10993$，　$\sum_i\sum_j y_{ik.}^2=32859$

$\sum_i y_{i..}^2=97687$，　$\sum_j y_{.j.}^2=130581$，　$\dfrac{(\sum\sum\sum y_{ijk})^2}{n}=10850.69$

$$S_A=\frac{1}{9}\times 97687-10850.69=3.42$$

$$S_B=\frac{1}{12}\times 130581-10850.69=31.06$$

$$S_{A\times B}=\frac{1}{3}\times 32859-10850.69-3.42-31.06=67.83$$

$$S_T=10993-10850.69=142.31$$

$$S_e=142.31-3.42-31.06-67.83=40$$

方差分析表

来源	平方和	自由度	均方和	F 比
机器 A	3.42	3	1.14	<1
操作工 B	31.06	2	15.53	9.30
交互作用 $A\times B$	67.83	6	11.305	6.77
e	40	24	1.67	
总和	1442.31	35	$F_{0.95}(2,24)=3.40$ $F_{0.95}(6,24)=2.51$	

由于 $F_B=9.30>3.40, F_{A\times B}=6.77>2.51$，所以在 $\alpha=0.05$ 水平上，操作工有显著差异，机器之间无显著差异，交互作用有显著差异.

3. 将抗生素注入人体会产生抗生素与血浆蛋白质结合的现象，以致减少了药效. 下表列出 5 种常用的抗生素注入牛的体内时，抗生素与血浆蛋白质结合的百分比.

青霉素	四环素	链霉素	红霉素	氯霉素

29.6	27.3	5.8	21.6	29.2
24.3	32.6	6.2	17.4	32.8
28.5	30.8	11.0	18.3	25.0
32.0	34.8	8.3	19.0	24.2

试在显著性水平$\alpha=0.05$下检验这些百分比的均值有无显著性的差异.

解: X_j,, $j=1,2,3,4,5$分别表示五种抗生素与血浆蛋白质结合的百分比,且假设$X_j \sim N(\mu_j,\sigma^2)$. 本题需要检验

$H_0:\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$

列表计算如下:

	青霉素	四环素	链霉素	红霉素	氯霉素	
	29.6	27.3	5.8	21.6	29.2	
	24.3	32.6	6.2	17.4	32.8	
	28.5	30.8	11	18.3	25	
	32	34.8	8.3	19	24.2	
T.j	114.4	125.5	31.3	76.3	111.2	458.7
T.j^2	13087.36	15750.25	979.69	5821.69	12365.44	48004.43
T.j^2/nj	3271.84	3937.563	244.9225	1455.423	3091.36	12001.11
ΣXij^2	3302.9	3967.73	261.97	1465.21	3139.12	12136.93

由上表计算得

$$S_T=12136.93-\frac{458.7^2}{20}=1616.65,\ S_A=12001.11-\frac{458.7^2}{20}=1480.83$$

$$S_E=1616.65-1480.83=135.82$$

S_T,S_A,S_E自由度依次为$n-1=19,s-1=4,n-s=15$,得方差分析表如下:

方差来源	平方和	自由度	均方和	F值
因素	1480.83	4	370.21	40.91
误 差	135.82	15	9.05	
总 和	1616.65	19		

若取$\alpha=0.05$,则$F_{0.05}(4,15)=3.06$,由于$F=40.91>F_{0.05}(4,15)$,故拒绝H_0,即认为五种抗生素与血浆蛋白质结合的百分比的均值有显著差异.

4. 今有不同温度处理的鱼卵胚胎发育速度（从受精到孵化所需时间）数据如下，试作方差分析（$\alpha = 0.01$）.

处理温度	胚胎发育速度数据				
21℃	128	129	132	130	134
23℃	123	125	126	127	128
25℃	99	100	102	110	105
27℃	86	88	90	93	95
29℃	76	75	78	80	81

解: X_j,, $j=1,2,3,4,5$ 分别表示五种温度处理的鱼卵胚胎发育速度，且假设 $X_j \sim N(\mu_j,\sigma^2)$. 本题需要检验

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

列表计算如下:

处理温度	胚胎发育速度数据					T. j	T. j^2	T. j^2/nj	Σ Xij^2
21℃	128	129	132	130	134	653	426409	85281. 8	85305
23℃	123	125	126	127	128	629	395641	79128. 2	79143
25℃	99	100	102	110	105	516	266256	53251. 2	53330
27℃	86	88	90	93	95	452	204304	40860. 8	40914
29℃	76	75	78	80	81	390	152100	30420	30446
合计						2640	1444710	288942	289138

由上表计算得:

$S_T = 289138 - \dfrac{2640^2}{25} = 10354$，$S_A = 288942 - \dfrac{2640^2}{25} = 10158$

$S_E = S_T - S_A = 10354 - 10158 = 196$

S_T, S_A, S_E 自由度依次为 $n-1=24, s-1=4, n-s=20$，得方差分析表如下:

方差来源	平方和	自由度	均方和	F 值
因素	10158	4	2539. 5	259. 1
误 差	196	20	9. 8	

总 和	10354	24		

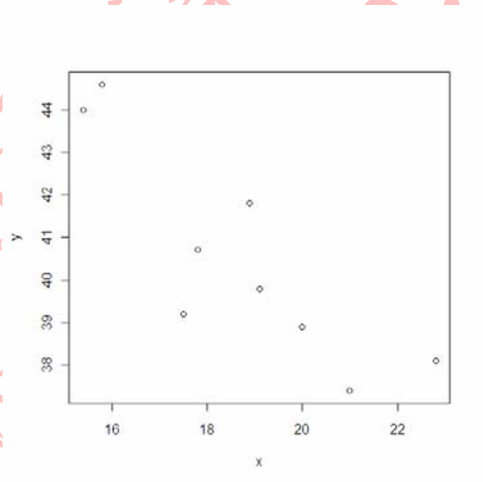
若取 $\alpha=0.01$，则 $F_{0.01}(4,20)=4.43$，由于 $F=259.1>F_{0.01}(4,20)$，故拒绝 H_0，即可认为不同温度处理的鱼卵胚胎发育速度具有显著差异.

5. 为了研究大豆脂肪含量（x）和蛋白质含量（Y）的关系，测定了 9 种大豆品种籽粒内的脂肪含量和蛋白质含量，得到如下表的数据:

品种编号	1	2	3	4	5	6	7	8	9
脂肪含量（%）x_i	15.4	17.5	18.9	20.0	21.0	22.8	15.8	17.8	19.1
蛋白质含量（%）y_i	44.0	39.2	41.8	38.9	37.4	38.1	44.6	40.7	39.8

试求出 Y 与 x 的关系.

解: 描图 Y 与 x 的散点图:



从散点图可以看出，Y 与 x 大致呈线性关系，现在 $n=9$，为求线性回归方程，所需计算列表如下:

	x	Y	x^2	Y^2	xY
	15.4	44	237.16	1936	677.6
	17.5	39.2	306.25	1536.64	686
	18.9	41.8	357.21	1747.24	790.02
	20	38.9	400	1513.21	778
	21	37.4	441	1398.76	785.4
	22.8	38.1	519.84	1451.61	868.68
	15.8	44.6	249.64	1989.16	704.68
	17.8	40.7	316.84	1656.49	724.46
	19.1	39.8	364.81	1584.04	760.18
总和	168.3	364.5	3192.75	14813.15	6775.02

由上表得

$$S_{xx}=3192.75-\frac{1}{9}168.3^2=45.54\text{，}S_{YY}=14813.15-\frac{1}{9}364.5^2=50.9$$

$$S_{xY}=6775.02-\frac{1}{9}168.3\times364.5=-41.13$$

故得 $\hat{\beta}_1=\frac{S_{xY}}{S_{xx}}=-\frac{41.13}{45.54}=-0.903$，

$$\hat{\beta}_0=\overline{Y}-\hat{\beta}_1\overline{x}=\frac{1}{9}\times364.5-(-0.903)\times\frac{1}{9}\times168.3=57.4$$

$$\hat{\sigma}^2=\frac{Q_\varepsilon}{n-2}=\frac{1}{n-2}\left[S_{YY}-\hat{\beta}_1S_{xY}\right]=\frac{1}{7}\left[50.9-(-0.903)\times(-41.13)\right]=1.966$$

于是得到回归方程

$$\hat{Y}=57.4-0.903x$$

下面我们对模型进行检验:

$$H_0:\beta_1=0\text{，}\quad H_1:\beta_1\neq0$$

H_0: $\beta_1=0$ 的拒绝域为

$$|t|=\frac{|\hat{\beta}_1|}{\hat{\sigma}}\sqrt{S_{xx}}\geq t_{\alpha/2}(n-2)$$

经查表 $t_{0.005}(7)=3.499$（$\alpha=0.01$），并把数据代入上式左边得

$$|t|=\frac{|\hat{\beta}_1|}{\hat{\sigma}}\sqrt{S_{xx}}=\frac{0.903}{\sqrt{1.966}}\sqrt{50.9}\approx4.595>3.499$$

故拒绝原假设 H_0: $\beta_1=0$，即可认为回归效果是显著的.

6. 对不同的元麦堆测得如下数据:

堆　号	1	2	3	4	5	6
重量 p	2813	2705	11103	2590	2131	5181
跨度 l	3.25	3.20	5.07	3.14	2.90	4.02

试求重量对跨度的回归方程，并求出根方差 σ 的估计值.

解: 设所求回归方程为 $\hat{p}=\hat{\beta}_0+\hat{\beta}_1 l$，由数据可以求出:

$$\sum_\alpha p_\alpha=26523,\sum_\alpha p_\alpha l_\alpha=109230.58,\sum_\alpha p_\alpha^2=176598625$$

$$\sum_{\alpha} l_{\alpha} = 21.58, \sum_{\alpha} l_{\alpha}^2 = 80.9374, N = 6$$

由最小二乘法估计公式可知

$$\hat{\beta}_1 = \frac{\sum_{\alpha} p_{\alpha} l_{\alpha} - \frac{1}{N} p_{\alpha} \sum_{\alpha} l_{\alpha}}{\sum_{\alpha} l_{\alpha}^2 - \frac{1}{N}\left(\sum_{\alpha} l_{\alpha}\right)^2} = 4165.85$$

$$\hat{\beta}_0 = \frac{1}{N}\sum_{\alpha} p_{\alpha} - \hat{\beta}_1 \frac{1}{N}\sum_{\alpha} l_{\alpha} = -10562$$

故可得回归方程： $\hat{p} = -10562 + 4165.85l$

σ^2 的估计是

$$\hat{\sigma}^2 = \frac{1}{N-2}\left\{\left[\sum_{\alpha} p_{\alpha}^2 - \frac{1}{N}(\sum_{\alpha} p_{\alpha})^2\right] - \hat{\beta}_1\left[\sum_{\alpha} p_{\alpha} l_{\alpha} - \frac{1}{N}(\sum_{\alpha} p_{\alpha})(\sum_{\alpha} l_{\alpha})\right]\right\}$$
$$= 428538$$

则 σ 的估计为 655

7. 某医院用光色比色计检验尿贡时，得尿贡含量与肖光系数读数的结果如下：

尿贡含量 x	2	4	6	8	10
肖光系数 y	64	138	205	285	360

已知它们之间有下述关系式：

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ i = 1,2,3,4,5.$$

各 ε_i 相互独立，均服从 $N(0,\sigma^2)$ 分布，试求 β_0, β_1 的最小二乘估计，并给出检验假设

$$H_0 : \beta_1 = 0$$

的拒绝域（ $\alpha = 0.01$ ）.

解：由数据可以求得，n=5

$$\sum_{\alpha} x_{\alpha} = 30，\ \bar{x} = 6$$

$$\sum_{\alpha} y_{\alpha} = 1052, \bar{y} = 210.4$$

$$\sum_{\alpha} x_{\alpha}^2 = 220, \sum_{\alpha} x_{\alpha} y_{\alpha} = 7790, \sum_{\alpha} y_{\alpha}^2 = 275990$$

$$l_{xx} = 40, l_{xy} = 1478, l_{yy} = 54649.2$$

则，最小二乘估计为：

$$\hat{\beta}_0 = -11.3, \hat{\beta}_1 = 36.95$$

检验假设 $H_0: \beta_1 = 0$ 可用统计量

$$F = \frac{\hat{\beta}_1 l_{xy}}{(l_{yy} - \hat{\beta}_1 l_{xy})/(n-2)} = 4416 > F_{1-\alpha}(1,3) = 34.1, \alpha = 0.01$$

因此，拒绝原假设，即可认为线性回归模型具有显著意义.

8. 研究同一地区土壤中所含植物可给态磷的情况，得到 18 组数据如下，其中，

x_1——土壤内所含无机磷浓度，

x_2——土壤内溶于 K2CO3 溶液并受溴化物水解的有机磷浓度，

x_3——土壤内溶于 K2CO3 溶液但不溶于溴化物的有机磷浓度，

y——载在 $20^{\circ}C$ 土壤内的玉米中可给态磷的浓度.

已知 y 与 x_1, x_2, x_3 之间有下述关系：

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \quad i = 1, 2, \cdots, 18$$

各 ε_i 相互独立，均服从 $N(0, \sigma^2)$ 分布，试求出回归方程.

土壤样本	x_1	x_2	x_3	y
1	0.4	53	158	64
2	0.4	23	163	60
3	3.1	19	37	71
4	0.6	34	157	61
5	4.7	24	59	54
6	1.7	65	123	77
7	9.4	44	46	81

8	10.1	31	117	93
9	11.6	29	173	93
10	12.6	58	112	51
11	10.9	37	111	76
12	23.1	46	114	96
13	23.1	50	134	77
14	21.6	44	73	93
15	23.1	56	168	95
16	1.9	36	143	54
17	26.8	28	202	168
18	29.9	51	124	99

解：由上述数据可以求得下面的结果：

$$p=3, n=18$$

$$\bar{x}_1=11.94, \bar{x}_2=42.11, \bar{x}_3=123, \bar{y}=81.28$$

$$L=\begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix}=\begin{pmatrix} 1752.9644 & 1085.6111 & 1200.0000 \\ 1085.6111 & 3155.7778 & 3364.0000 \\ 1200.0000 & 3364.0000 & 35572.0000 \end{pmatrix}$$

$$l=\begin{pmatrix} l_{10} \\ l_{20} \\ l_{30} \end{pmatrix}=\begin{pmatrix} 3231.4778 \\ 2216.4445 \\ 7953.0000 \end{pmatrix}$$

$$L^{-1}=\begin{pmatrix} l^{11} & l^{12} & l^{13} \\ l^{21} & l^{22} & l^{23} \\ l^{31} & l^{32} & l^{33} \end{pmatrix}=\begin{pmatrix} 0.000725 & -0.000248 & -0.000001 \\ -0.000248 & 0.000437 & -0.000033 \\ -0.000001 & -0.000033 & 0.000031 \end{pmatrix}$$

$$\hat{\beta}=\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}=L^{-1}l=\begin{pmatrix} 1.784780 \\ -0.083397 \\ 0.161133 \end{pmatrix}$$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}_1-\hat{\beta}_2\bar{x}_2-\hat{\beta}_3\bar{x}_3=43.652198$$

所求得的回归方程为

$$\hat{y}=43.65+1.78x_1-0.08x_2+0.16x_3$$

记

$$S_T=\sum_{\alpha}(y_\alpha-\bar{y})^2=12389.6111$$

$$S_R=\sum_{j=1}^{3}\hat{\beta}_j l_{j0}=6806.1115$$

$$S_e=S_T-S_R=5583.4997$$

对方乘作检验的 F 统计量为:

$$F=\frac{S_R/p}{S_e/(n-p-1)}=5.6885>F_{1-0.05}(3,14)=3.34$$

故在 $\alpha=0.05$ 的水平上方程是显著的.

对各因子作 F 检验的统计量分别为

$$F_1=\frac{\hat{\beta}_1^2}{l^{11}S_e/(n-p-1)}=11.02>F_{0.95}(1,14)=4.60$$

$$F_2=\frac{\hat{\beta}_2^2}{l^{22}S_e/(n-p-1)}=0.0399<F_{0.95}(1,14)=4.60$$

$$F_3=\frac{\hat{\beta}_3^2}{l^{33}S_e/(n-p-1)}=2.0822<F_{0.95}(1,14)=4.60$$

故在 $\alpha=0.05$ 的水平上, x_1 是显著的, x_2 与 x_3 是不显著的.

习题 B

1. 在某橡胶配方中，考虑了 3 种不同的促进剂，4 种不同分量的氧化锌，每种配方各做一次试验，测得 300%定强如下表:

促进剂 A	氧化锌 B			
	B_1	B_2	B_3	B_4
A_1	31	34	35	39
A_2	33	36	37	38
A_3	35	37	39	42

试问促进剂、氧化剂对定强有无显著影响?

解:本题是两因素无重复试验的方差分析问题。列出方差计算表如下:

因素 B	因素 A			$\sum_{j=1}^{3} x_{ij}$	$\left(\sum_{j=1}^{3} x_{ij}\right)^2$
	A1	A2	A3		
B1	31	33	35	99	9801
B2	34	36	37	107	11449
B3	35	37	39	111	12321
B4	39	38	42	119	14161
$\sum_{i=1}^{4} x_{ij}$	139	144	153	436	47732
$\left(\sum_{i=1}^{4} x_{ij}\right)^2$	19321	20736	23409	63446	
$\sum_{i=1}^{4} \left(x_{ij}\right)^2$	4863	5198	5879	15940	

由上表算出:

$$P=\frac{1}{lm}\left(\sum_{i=1}^{l}\sum_{j=1}^{m} x_{ij}\right)^2=\frac{1}{4\times 3}\times 436^2=15841.33$$

$$Q_A=\frac{1}{l}\sum_{j=1}^{m}\left(\sum_{i=1}^{l} x_{ij}\right)^2=\frac{1}{4}\times 63466=15866.5$$

$$Q_B=\frac{1}{m}\sum_{i=1}^{l}\left(\sum_{j=1}^{m} x_{ij}\right)^2=\frac{1}{3}\times 47732=15910.67$$

$$R=\frac{1}{m}\sum_{i=1}^{l}\sum_{j=1}^{m}\left(x_{ij}\right)^2=\frac{1}{3}\times 47732=15910.67$$

$$S_A^2=Q_A-P=15866.5-15841.33=25.17$$

$$S_B^2=Q_B-P=15910.67-15841.33=69.34$$

$$S_T^2=R-P=15940-15841.33=98.67$$

$$S_E^2=S_T^2-S_A^2-S_B^2=98.67-25.17-69.34=4.16$$

列出方差分析表:

方差来源	平方和	自由度	下值	显著性
A	25.17	2	$F_A = 18.16$	**
B	69.43	3	$F_B = 33.35$	**
误差	4.16	6		
总和	98.67	11		

查 F 分布表得，$F_{0.05}(2,6)=5.14$，$F_{0.05}(3,6)=4.76$，$F_{0.01}(2,6)=10.94$，$F_{0.01}(3,6)=9.78$。因 $F_A=18.16>10.94$，$F_B=33.35>9.78$，所以促进剂、氧化锌对定强的影响都是高度显著的。

2. 设

$$
\begin{aligned}
&y_i=\beta_0+\beta_1 x_i+\beta_2(3x_i^2-2)+\varepsilon_i \quad i=1,2,3,\\
&x_1=-1, x_2=0, x_3=1.
\end{aligned}
$$

$\varepsilon_1,\varepsilon_2,\varepsilon_3$ 相互独立同服从于 $N(0,\sigma^2)$.

(1) 写出矩阵 X；

(2) 求 β_0,β_1,β_2 的最小二乘估计;

(3) 证明当 $\beta_2=0$ 时，β_0,β_1 的最小二乘估计不变.

解：（1）$X=\begin{pmatrix}1 & -1 & 1\\ 1 & 0 & -2\\ 1 & 1 & 1\end{pmatrix}$

（2）$X'X=\begin{pmatrix}3 & 0 & 0\\ 1 & 2 & 0\\ 0 & 0 & 6\end{pmatrix}$，$X'Y=\begin{pmatrix}y_1+y_2+y_3\\ -y_1+y_3\\ y_1-2y_2+y_3\end{pmatrix}$，则，$\beta_0,\beta_1,\beta_2$ 的最小二乘估计是

$$
\hat\beta=\begin{pmatrix}\hat\beta_0\\ \hat\beta_1\\ \hat\beta_2\end{pmatrix}=(X'X)^{-1}X'Y=\begin{pmatrix}\frac{1}{3}(y_1+y_2+y_3)\\ \frac{1}{2}(-y_1+y_3)\\ \frac{1}{6}(y_1-2y_2+y_3)\end{pmatrix}
$$

（3）若 $\beta_2=0$，此时模型成为：

$y_i=\beta_0+\beta_1 x_i+\varepsilon_i \quad i=1,2,3$，则对应的

$X=\begin{pmatrix}1 & -1\\ 1 & 0\\ 1 & 1\end{pmatrix}$，$X'X=\begin{pmatrix}3 & 0\\ 0 & 2\end{pmatrix}$，$X'Y=\begin{pmatrix}y_1+y_2+y_3\\ -y_1+y_3\end{pmatrix}$，$\beta_0,\beta_1$ 的最小二乘估计是

$$\hat{\beta}=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}=(X'X)^{-1}X'Y=\begin{pmatrix}\frac{1}{3}(y_1+y_2+y_3)\\ \frac{1}{2}(-y_1+y_3)\end{pmatrix}$$

3. 若 y 与 x 有下述关系：

$$y=\beta_0+\beta_1x+\beta_2x^2+\cdots+\beta_px^p+\varepsilon.$$

其中 $\varepsilon\sim N(0,\sigma^2)$ 从中获得了 n 组独立观测值 (x_α,y_α)，能否求出 $\beta_0,\beta_1,\cdots,\beta_p$ 的最小二乘估计，试写出最小二乘估计的公式.

解：若记 $X_{\alpha i}=x_\alpha^i,\bar{X}_i=\frac{1}{n}\sum_{\alpha=1}^{n}x_\alpha^i\quad \alpha=1,\cdots,n;i=1,\cdots,p$

$$l_{ij}=\sum_{\alpha=1}^{n}(X_{\alpha i}-\bar{X}_i)(X_{\alpha j}-\bar{X}_j)\ \ i,j=1,\cdots,p$$

$$l_{i0}=\sum_{\alpha=1}^{n}(X_{\alpha i}-\bar{X}_i)(y_\alpha-\bar{y})\ \ i=1,\cdots,p$$

则 $\beta_1,\cdots,\beta_p$ 的最小二乘估计为下述方程组的解：

$$\begin{cases}l_{11}\hat{\beta}_1+l_{12}\hat{\beta}_2+\cdots+l_{1p}\hat{\beta}_p=l_{10}\\ l_{21}\hat{\beta}_1+l_{22}\hat{\beta}_2+\cdots+l_{2p}\hat{\beta}_p=l_{20}\\ \qquad\cdots\qquad\cdots\\ l_{p1}\hat{\beta}_1+l_{p2}\hat{\beta}_2+\cdots+l_{pp}\hat{\beta}_p=l_{p0}\end{cases}\qquad (*)$$

β_0 的最小二乘估计为：

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{X}_1-\cdots-\hat{\beta}_p\bar{X}_p$$

若把方程组（*）的系数矩阵记为 L，则 $L=(l_{ij})$，又记 $L^{-1}=(l^{ij})$，则在显著性水平 α 上检验 $H_0:\beta_i=0$ 的拒绝域是：

$$F_i=\frac{\hat{\beta}_i^2}{l^{ii}\hat{\sigma}^2}>F_{1-\alpha}(1,n-p-1)$$

其中，$\hat{\sigma}^2 = \frac{1}{n-p-1}\{\sum_{\alpha}(y_\alpha - \bar{y})^2 - \hat{\beta}_1 l_{10} - \cdots - \hat{\beta}_p l_{p0}\}$

4．研究货运总量 y（万 t）与工业总产值 x_1（亿元）、农业总产值 x_2（亿元）、居民非商品支出 x_3（亿元）的关系，其数据见下表，试求 y 关于 x_1，x_2，x_3 的三元线性回归方程.

编号	货运总量 y/亿吨	工业总产值 x_2/亿元	农业总产值 x_3/亿元	居民非商品支出 x_3/亿元
1	160	70	35	1.0
2	260	75	40	2.4
3	210	65	40	2.0
4	265	74	42	3.0
5	240	72	38	1.2
6	220	68	45	1.5
7	275	78	42	4.0
8	160	66	36	2.0
9	275	70	44	3.2
10	250	65	42	3.0

解:

$$X = \begin{pmatrix} 1 & 70 & 35 & 1.0 \\ 1 & 75 & 40 & 2.4 \\ 1 & 65 & 40 & 2.0 \\ 1 & 74 & 42 & 3.0 \\ 1 & 72 & 38 & 1.2 \\ 1 & 68 & 45 & 1.5 \\ 1 & 78 & 42 & 4.0 \\ 1 & 66 & 36 & 2.0 \\ 1 & 70 & 44 & 3.2 \\ 1 & 65 & 42 & 3.0 \end{pmatrix}, Y = \begin{pmatrix} 160 \\ 260 \\ 210 \\ 265 \\ 240 \\ 220 \\ 275 \\ 160 \\ 275 \\ 250 \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

经计算 $\widehat{\beta}=\left(X^{T}X\right)^{-1}X^{T}Y=\begin{pmatrix}-348.280\\3.754\\7.101\\12.447\end{pmatrix}$

于是得到回归方程为

$$\hat{Y}=-348.280+3.754x_1+7.101x_2+12.447x_3$$

与一元线性回归方法一样,为了考察模型的结构是否符合实际观察结果,还需进行以下的假设检验:

$$H_0:\beta_0=\beta_1=\cdots=\beta_p=0,$$

$$H_1:\beta_i\text{不全为零}.$$

若在显著性水平α下拒绝H_0. 我们就认为回归效果是显著的. (具体检验过程略)