

第一章 概率

一、事件

- 1、必然事件:在一定的条件下必然要发生的事件叫必然事件。
- 2、不可能事件:在一定的条件下不可能发生的事件叫不可能事件。
- 3、随机事件:在一定的条件下可能发生也可能不发生的事件叫随机事件。

*事件的两要素:一条件 (或试验), 二结果

口答: 指出下列事件是必然事件, 不可能事件, 还是随机事件:

- (1) 如果 a, b 都是实数, 那么 $a \cdot b = b \cdot a$ 。
- (2) 从一批兵乓球中抽出一个, 是优等品
- (3) 投掷一枚硬币, 出现正面
- (4) 投掷一个骰子, 出现 3 点
- (5) 投掷一个骰子, 数点大于 2
- (6) 投掷一个骰子, 数点大于 6

二、事件的频率

在 n 次重复同一试验时, 事件 A 发生的次数为 n_A 次, 则称 $\frac{n_A}{n}$ 为事件 A 发

生的频率记作, $f_n(A) = \frac{n_A}{n}$, 注: $0 \leq f(A) \leq 1$

(引例、设事件 $A = \{\text{抛掷一枚硬币, 出现正面}\}$, 这是一个随机事件, 抛掷一枚硬币这个试验可以重复进行, 当抛掷一枚硬币 10 次时, 有 6 次出现正面, 也就是说在 10 次试验中事件 A 发生了 6 次, 以后我们把事件 A 发生的次数与试验的次数的比值叫做事件 A 发生的频率)

三、事件的概率

当试验的次数很多很多时, 事件的频率总是接近于某个常数, 在他的附近摆动, 统计学家们发现只要是他们所研究的事件都有这样的一个常数, 可见这是一个一般的规律, 以后我们把这个常数叫做事件的概率。事件 A 的概率可记作 $P(A)$

当 n 很大时 A 发生的频率 $f_n(A)$ 接近于某个常数 p , 则常数 p 叫做事件 A 的概率, 记作 $P(A)=p$ 。

四、概率的性质

(1) $0 \leq P(A) \leq 1$

(2) 必然事件的概率为 1, 不可能事件的概率为 0。反之不成立。

(3) 当事件 A 确定时, 频率 $f_n(A)$ 试验的次数有关, 概率 $P(A)$ 与试验的次数无关是固定的, 但我们可以用频率来估计概率。

五、事件的关系与运算

1、包含关系: 若 A 发生则 B 一定发生, 则 $A \subseteq B (B \supseteq A)$

2、相等关系: 若 $A \subseteq B, B \subseteq A$, 则 $A = B$

3、和事件

$A+B = \{A \text{ 与 } B \text{ 至少有一个发生}\}$, 也可记作 $A \cup B$

例 2、投掷一个骰子,

$A = \{\text{点数} < 4\}$, $B = \{\text{偶数点}\}$, 则 $A+B =$ _____

4、积事件的积 (交)

$AB = \{A \text{ 与 } B \text{ 同时发生}\}$, 也可记作 $A \cap B$

例 3、投掷一个骰子,

$A = \{\text{点数} < 4\}$, $B = \{\text{奇数点}\}$, 则 $AB =$ _____

5、互斥事件: 若 $A \cap B = \phi$, 则 A 与 B 叫做互斥事件

例 4、投掷一个骰子,

$A = \{\text{点数} < 3\}$, $B = \{\text{偶数点}\}$, $C = \{\text{点数} > 4\}$

6、对立事件: 若 A 与 B 是互斥事件, 且 $A+B$ 是必然事件, 则 A 与 B 叫做对立事件, 记作 $B = \bar{A}$ 或 $A = \bar{B}$

7、公式:

(1) 若 A 与 B 是互斥事件, 则 $P(A+B) = P(A) + P(B)$

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

例 1、从一袋子中装有红白黄三色的球, 从中取一球是红球的概率为 $\frac{1}{2}$, 是白球的概率为 $\frac{1}{6}$

求: (1) 从中取一球是红球或白球的概率

(2) 从中取一球是黄球的概率

例 2、高三 800 名学生的质检数学成绩中抽取一个样本频率分布直方图如图. 图中成绩在 125~140 分的频数是 6.

(1) 求样本容量; (2) 成绩在 80~95 分之间的总人数 (3) 估计成绩在 80~110 分概率

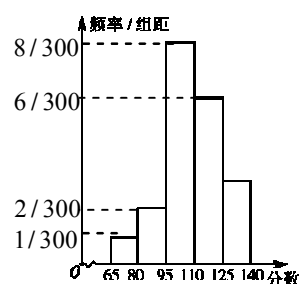
解: (1) 设样本容量为 n, 因成绩在 125~140 分的频率为 $1 - \frac{1}{20} - \frac{2}{20} - \frac{8}{20} - \frac{6}{20} = \frac{3}{20}$,

于是 $\frac{6}{n} = \frac{3}{20}$, $n = 40$

(2) $800 \times \frac{2}{20} = 80$, 于是成绩在 80~95 分之间的总人数约为 80 人

(3) 因成绩在 80~110 分的频率为 $\frac{2+8}{20} = \frac{1}{2}$

于是成绩在 80~110 分概率约为 0.5



二、古典概型

引例 1: 投一个骰子有六种可能的结果, 每种结果是等可能的, 这里的每种结果叫基本事件, 于投一个骰子有 6 个基本事件, 设 $A = \{\text{点数} < 3\}$, A 含有 2 个基本事件,

于是 $P(A) = \frac{2}{6}$

引例 2: 投 1 次硬币有正反面两种结果, 投 2 次硬币有几种结果呢? 4 个。

它们是正正, 正反, 反正, 反反, 有 4 个基本事件, 设 $A = \{\text{至少出现一次正面}\}$

则 A 含有 3 个基本事件, $P(A) = \frac{3}{4}$

1、基本事件: 设一个试验有 n 个等可能的结果, 每个结果就叫做此试验的基本事件, 每个基本事件的概率都是 $\frac{1}{n}$

2、古典概型: 设一个试验有 n 个基本事件, 若事件 A 含有 n_A 个基本事件, 则事件 A 的概率 $P(A) = \frac{n_A}{n}$, 这样概率问题叫做古典概型。

例 1、投掷 2 次骰子, 求

(1) 点数和为 5 的概率

(2) 第 1 次为偶数点, 第 2 次点数大于 2 点的概率

思考: 将骰子抛掷 2 次, 求每次都是奇数点的概率

例 2、一袋中有红球 4 个，白球 2 个，求下列事件的概率

- (1) 每次取 1 球，取后放回，取 2 次都取到红球
- (2) 每次取 1 球，取后不放回，取 2 次都取到红球
- (3) 从中取 2 球，取得 2 球都是红球

练习

1、连掷两次骰子得到的点数分别为 m 和 n ，求 $m > n$ 的概率

2、若 $a \in \{0, 1, 2, 3\}$ ， $b \in \{0, 1, 2\}$ ，求方程 $x^2 + 2ax + b^2 = 0$ 有实根的概率。

3、箱中装有 15 张卡片，每张正面分别标有 1 到 15 中的一个号码，正面号码为 n 的卡片反面标的数字是 $n^2 - 12n + 40$

(I) 任意取出一张卡，求正面数字大于反面数字的概率；

(II) 同时取出两张卡片，求他们反面数字相同的概率。

解：(1) 由不等式 $n > n^2 - 12n + 40$ ，得 $5 < n < 8$ 。……………2 分

由题意得， $n = 6, 7$ 。所求的概率为 $\frac{2}{15}$ 答：所求的概率为 $\frac{2}{15}$

(2) 设取出的是第 m 号卡片和 n 号卡片 ($m \neq n$)，则有

$m^2 - 12m + 40 = n^2 - 12n + 40$ 即 $12(n - m) = n^2 - m^2$ ，由 $m \neq n$ 得 $m + n = 12$

符合条件的取法为 1, 11; 2, 10; 3, 9; 4, 8; 5, 7; 故所求的概率为 $\frac{5}{C_{15}^2} = \frac{1}{21}$

4. 一次抽奖活动规则是：从一个装有分值分别为 1, 2, 3, 4, 5, 6 的六个相同小球的抽奖箱中，有放回的抽取两次，每次抽取一个球，设分值和为 a 。规定：当 $a = 12$ 时，则获一等奖，当 $a = 10$ 或 11 时，则获二等奖，当 $a < 10$ 时则无；

(1) 获一等奖的概率；(2) 获二等奖的概率 (3) 获奖的概率 (4) 不获奖的概率

解：(1) 设获一等奖为事件 A，获二等为事件 B

则所有的取球方法有

(1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)

(2, 1) (2, 2) (2, 3) (2, 4) (2, 5) (2, 6)

……

(6, 1) (6, 2) (6, 3) (6, 4) (6, 5) (6, 6)

于是基本事件总数 $n = 36$ ， $n_A = 1$ ，故 $P(A) = \frac{1}{36}$

(2) $n_B = 5$ ， $P(B) = \frac{5}{36}$

(3) 因为 $A \cap B = \emptyset$ ，所以获奖的概率 $= P(A + B) = P(A) + P(B) = \frac{1}{36} + \frac{5}{36} = \frac{1}{6}$

(4) 不获奖的概率 $= 1 - \frac{1}{6} = \frac{5}{6}$

5、甲有一个装有 3 个红球、3 个黑球的箱子，乙有一个装有 3 个红球、2 个黑球的箱子，两人各自从自己的箱子里任取一球，并约定：所取两球同色时甲胜，异色时乙胜求甲获胜的概率；

解：设甲、乙都取红球为事件 A，甲、乙都取黑球为事件 B

甲的箱子的 3 个红球与 3 个黑球分别记为 $a_1, a_2, a_3, a_4, a_5, a_6$

乙的箱子的 3 个红球与 2 个黑球分别记为 b_1, b_2, b_3, b_4, b_5

则所有的取球方法有

(a_1, b_1) (a_1, b_2) (a_1, b_3) (a_1, b_4) (a_1, b_5)

(a_2, b_1) (a_2, b_2) (a_2, b_3) (a_2, b_4) (a_2, b_5)

……

(a_6, b_1) (a_6, b_2) (a_6, b_3) (a_6, b_4) (a_6, b_5)

于是 $n=30, n_A=9, n_B=6$

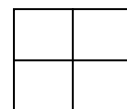
$$\text{故 } P(A) = \frac{9}{30} = \frac{3}{10}, \quad P(B) = \frac{6}{30} = \frac{1}{5}$$

$$\therefore \text{甲获胜的概率 } P = P_1 + P_2 = \frac{3}{10} + \frac{1}{5} = \frac{1}{2}$$

三、几何概型

引例 1、如图在正方形内任取一点，正方形内每个点被取到的机会相等，求在所取的点在左上角的正方形内的概率。

解：设 $A = \{\text{在所取的点在左上角的正方形内}\}$



试验点组成的图形面积为 $S_\Omega = a^2$

$$\text{事件点组成的图形的面积为 } S_A = \left(\frac{a}{2}\right)^2 \quad \text{则 } P(A) = \frac{S_A}{S_\Omega} = \frac{1}{4} \text{ 答}$$

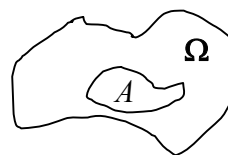
引例 2、取一根长度为 $3m$ 的绳子，拉直后在任意位置剪断，剪得两段的长都不小于 $1m$ 的概率有多大？

$$\text{解：所求事件为 } A, \quad l_\Omega = 3, \quad l_A = 1, \quad \text{则 } P(A) = \frac{l_A}{l_\Omega} = \frac{1}{3}$$

引例 3、在正方体内任取一点，求这点在内切球内的的概率

1、几何概型：在几何区域 D 中随机地取一点，该区域中每一点被取到的机会相同，解决这种试验下的事件的概率模型叫几何概型

$$2、P(A) = \frac{S_A}{S_\Omega} (\text{或 } \frac{l_A}{l_\Omega} \text{ 或 } \frac{V_A}{V_\Omega})$$



例 1、在正三角形内任取一点，求这点在内切圆内的的概率

例 2、A 是圆上的定点，在圆上任取一点 P，求点 P 到 A 的距离小于半径的概率。

例 3、在区间 $[0,4]$ 上任取两个数 x 与 y ，求

(1) $x > 3$ 的概率 (2) $y \geq x + 2$ 的概率

解：(1) 试验点集 $\Omega_1 = \{x | 0 \leq x \leq 4\}$ ，事件点集 $A = \{x | 3 < x \leq 4\}$

$$\text{则 } l_\Omega = 4, \quad l_A = 1, \quad \text{则 } P(A) = \frac{l_A}{l_\Omega} = \frac{1}{4}$$

(2) 试验点集 $\Omega_2 = \{(x,y) | 0 \leq x \leq 4, 0 \leq y \leq 4\}$ ，

事件点集 $A = \{(x,y) | 0 \leq x \leq 4, 0 \leq y \leq 4, y \geq x + 2\}$

$$\text{则 } S_\Omega = 16, \quad S_A = 2, \quad \text{则 } P(A) = \frac{S_A}{S_\Omega} = \frac{2}{16} = \frac{1}{8}$$



例 4、送报人每人早上 6:30 到 7:30 之间把报纸送到老李家里，但老李每天早上 7:00 到

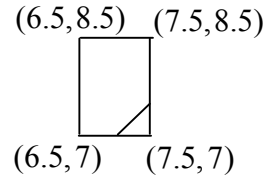
8: 30 之间离家, 求老李离家前能得到报纸的概率

解: 设送报人到老李家时刻为 x 点, 老李离家的时刻为 y 点

试验点集 $\Omega = \{(x, y) | 6.5 \leq x \leq 7.5, 7 \leq y \leq 8.5\}$,

事件点集 $A = \{(x, y) | 6.5 \leq x \leq 7.5, 7 \leq y \leq 8.5, y \geq x\}$

$$\text{则 } S_{\Omega} = \frac{3}{2}, \quad S_A = \frac{3}{2} - \frac{1}{8} = \frac{11}{8}, \quad \text{则 } P(A) = \frac{S_A}{S_{\Omega}} = \frac{\frac{11}{8}}{\frac{3}{2}} = \frac{11}{12}$$



例 5、在 $\triangle ABC$ 中, $\angle BAC = 90^\circ$, $\angle B = 30^\circ$, $AD \perp BC$ 于 D

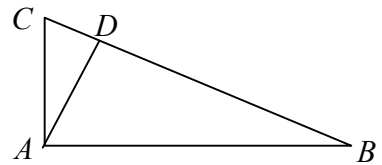
(1) 在 $\triangle ABC$ 内任取一点 N , 求 N 点在 $\triangle ACD$ 内的概率;

(2) 在 BC 上任取一点 M , 求 $\triangle AMB$ 为钝角三角形的概率;

(3) 在 $\angle BAC$ 内任作一条射线 AT , 求 $\angle BAT$ 小于 30° 的概率

解: (1) 设所求事件为 A , 则

试验点集 Ω 为 $\triangle ABC$ 区域, 事件点集 A 为 $\triangle ADC$ 区域



$$\text{设 } AC=1, \quad S_{\Omega} = S_{\triangle ABC} = \frac{\sqrt{3}}{2}, \quad S_A = S_{\triangle ADC} = \frac{\sqrt{3}}{8}, \quad \text{于是 } P(A) = \frac{S_A}{S_{\Omega}} = \frac{1}{4}$$

(2) 设所求事件为 B , 则

试验点集 Ω 为线段 BC , 事件点集 B 为线段 BD

$$l_{\Omega} = BC = 2, \quad l_B = BD = \frac{3}{2}, \quad \text{于是 } P(B) = \frac{l_B}{l_{\Omega}} = \frac{\frac{3}{2}}{2} = \frac{3}{4}$$

(3) 设 $\angle BAT = x$, 则

试验点集 $\Omega = \{x | 0 \leq x \leq \frac{\pi}{2}\}$, 事件点集 $A = \{x | 0 \leq x \leq \frac{\pi}{6}\}$

$$\text{于是 } l_{\Omega} = \frac{\pi}{2}, \quad l_A = \frac{\pi}{6}, \quad \text{于是 } P(A) = \frac{l_A}{l_{\Omega}} = \frac{\frac{\pi}{6}}{\frac{\pi}{2}} = \frac{1}{3}$$

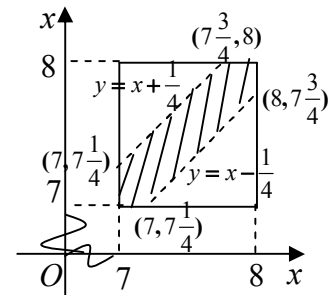
例 6、甲、乙两人约定在 7 时到 8 时之间在某处会面, 并约定先到者应等候另一个人一刻钟, 过时即可离去, 求两人能会面的概率.

设 $A = \{\text{能会面}\}$, 甲、乙两人到会面处的时刻是 x, y , 则

试验点集 $\Omega = \{(x, y) | 7 \leq x \leq 8, 7 \leq y \leq 8\}$

事件点集 $A = \{(x, y) | (x, y) \in \Omega, |x - y| \leq \frac{1}{4}\}$

$$|x - y| \leq \frac{1}{4} \Rightarrow -\frac{1}{4} \leq x - y \leq \frac{1}{4} \Rightarrow \begin{cases} x - y \leq \frac{1}{4} \\ x - y \geq -\frac{1}{4} \end{cases} \Rightarrow \begin{cases} y \geq x - \frac{1}{4} \\ y \leq x + \frac{1}{4} \end{cases}$$



作出图象如图

$$S_{\Omega}=1, S_A=\frac{7}{16}, P(A)=\frac{S_A}{S_{\Omega}}=\frac{\frac{7}{16}}{1}=\frac{7}{16}, \text{答: 两人能会面的概率是 } \frac{7}{16}$$

例 7、甲、乙两船驶向一个不能同时停泊两艘船的码头, 它们在一昼夜内到达码头的时刻是等可能的, 如果甲船停泊时间为 1h, 乙船停泊时间为 2h, 求它们中的任意一艘都不需要等待码头空出的概率.

解: 设 $A=\{\text{意一艘都不需要等待码头空出}\}$

甲、乙两人到码头的时刻是 x, y , 则

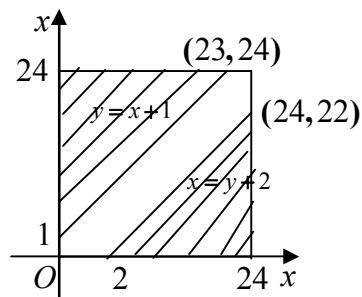
试验点集 $\Omega=\{(x,y)|0 \leq x \leq 24, 0 \leq y \leq 24\}$

事件点集 $A=\{(x,y)|x > y+2, \text{或 } y > x+1, (x,y) \in \Omega, \}$

作出图象如图

$$S_{\Omega}=24^2, S_A=\frac{1}{2} \times 22^2 + \frac{1}{2} \times 23^2, P(A)=\frac{S_A}{S_{\Omega}}=\frac{1013}{1152}$$

答: 所求概率是 $\frac{1013}{1152}$

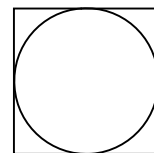


3、随机估计

例 5、在图的正方形中随机撒一把芝麻, 用随机模拟的方法来估计圆周率 π 的值. 如果撒了 1000 个芝麻, 落在圆内的芝麻总数是 776 颗, 那么这次模拟中 π 的估计值是_____. (精确到 0.001)

练习

练习:



1、小张的手表停了, 打开收音机想听报时, 则他等待的时间不多于 10 分钟的概率

2、在区间 $[-\frac{\pi}{2}, \frac{\pi}{2}]$ 上随机取一个数 x , $\cos x$ 的值介于 0 到 $\frac{1}{2}$ 之间的概率为(). A. $\frac{1}{3}$

B. $\frac{2}{\pi}$ C. $\frac{1}{2}$ D. $\frac{2}{3}$

3、在区间 $[-\pi, \pi]$ 内随机取两个数分别记为 a, b , 则使得函数 $f(x)=x^2+2ax-b^2+\pi^2$ 有零点的概率为_____

4、把一线段分成三段能组成三角形的概率

解: 设线段长 1, 三段长分别为 $x, y, 1-x-y$, 于是

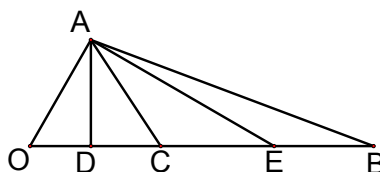
试验点集 $\Omega=\{(x,y)|x>0, y>0, x+y<1\}$

$$\text{组成三角形的条件是 } \begin{cases} x+y > 1-x-y \\ x+1-x-y > y \\ y+1-x-y > x \end{cases} \Rightarrow \begin{cases} x+y > \frac{1}{2} \\ y < \frac{1}{2} \\ x < \frac{1}{2} \end{cases}$$

事件点集 $A=\{(x,y)|x+y > \frac{1}{2}, y < \frac{1}{2}, x < \frac{1}{2}\}$

$$\text{则 } S_{\Omega}=\frac{1}{2}, S_A=\frac{1}{8}, \text{ 则 } P(A)=\frac{S_A}{S_{\Omega}}=\frac{\frac{1}{8}}{\frac{1}{2}}=\frac{1}{4}$$

例 5、在 $\triangle ABC$ 中, $\angle AOB=60^\circ$, $OA=2$, $OB=5$,



(1) 在线段 OB 上任取一点 C ,

求 $\triangle AOC$ 为钝角三角形的概率;

(2) 在 $\angle OAB$ 内任作一条射线 OT ,

求 $\angle OAT$ 小于 30° 的概率

解: 如图, 由平面几何知识:

当 $AD \perp OB$ 时, $OD = 1$;

当 $OA \perp AE$ 时, $OE = 4$, $BE = 1$.

(1) 当且仅当点 C 在线段 OD 或 BE 上时, $\triangle AOC$ 为钝角三角形

记 " $\triangle AOC$ 为钝角三角形" 为事件 M , 则 $P(M) = \frac{OD+EB}{OB} = \frac{1+1}{5} = 0.4$

即 $\triangle AOC$ 为钝角三角形的概率为 0.4.

(2) 当且仅当点 C 在线段 DE 上时, $\triangle AOC$ 为锐角三角,

记 " $\triangle AOC$ 为锐角三角" 为事件 N , 则 $P(N) = \frac{DE}{OB} = \frac{3}{5} = 0.6$

即 $\triangle AOC$ 为锐角三角形的概率为 0.6.

四、条件概率

1、条件概率的古典式 $P(B|A) = \frac{n(AB)}{n(A)}$

引例: 设投一枚骰子记事件 $A = \{\text{点数奇数}\}$, 事件 $B = \{\text{点数大于 1}\}$

则 $P(A) = \frac{1}{2}$, $P(B) = \frac{5}{6}$,

若知道事件 A 发生的条件下, 求事件 B 发生的概率。此概率记为 $P(B|A)$

在事件 A 发生的条件下, 基本事件只有 1 点, 3 点, 5 点, 即 $n(A) = 3$

其中使 B 发生的有 3 点, 5 点, $n(AB) = 2$, 于是 $P(B|A) = \frac{2}{3}$

例 1、设投一枚骰子记事件 $A = \{\text{点数小于 4}\}$, 事件 $B = \{\text{点数为偶数}\}$, 求 $P(B|A)$

解: $n(A) = 3, n(AB) = 1, P(B|A) = \frac{1}{3}$

例 2、一袋中有红球 3 个, 白球 2 个, 每次一球, 取后不放回, 求在第一次取红球的条件下, 第二次取白球的概率

解: 设 $A = \{\text{第一次取红球}\}, B = \{\text{第二次取白球}\}$

$n(A) = 3 \times 4, n(AB) = 3 \times 2, P(B|A) = \frac{3 \times 2}{3 \times 4} = \frac{1}{2}$ 另解: $P(B|A) = \frac{2}{4} = \frac{1}{2}$

2、条件概率的一般式 $P(B|A) = \frac{P(AB)}{P(A)}$, $P(AB) = P(A)P(B|A)$

例 1、一袋中有红球 3 个, 白球 2 个, 每次一球, 取后不放回, 求第一次取红球且第二次取白球的概率

例 2、一袋中有 3 个红球, 2 个白球, 每次取一球, 取后不放回, 当取到白球后就停止取球, 求

(1) 取二次停止的概率 (2) 取三次停止的概率

(3) 取不超过二次就停止的概率;

解: 设 $A_i = \{\text{第 } i \text{ 次取到白球}\}$

(1) $P(\overline{A_1}A_2) = P(\overline{A_1})P(A_2|\overline{A_1}) = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$

$$(2) P(\overline{A_1}\overline{A_2}A_3) = P(\overline{A_1})P(\overline{A_2}|\overline{A_1})P(A_3|\overline{A_1}\overline{A_2}) = \frac{3}{5} \times \frac{2}{4} \times \frac{2}{3} = \frac{2}{5}$$

(3) 不超过二次就停止记为事件 A

$$\text{则 } A = A_1 + \overline{A_1}A_2, P(A) = P(A_1) + P(\overline{A_1}A_2) = \frac{2}{5} + \frac{3}{10} = \frac{7}{10}$$

注: (1) 当 A 发生后的状况明确可用状况直接求 $P(B|A)$

$$(2) P(ABC) = P(AB)P(C|AB) = P(A)P(B|A)P(C|AB)$$

N 个事件积的概率等于第一事件的概率乘以第一事件发生的条件下第二事件的概率再乘以第一二事件都发生的条件下第三事件的概率...

例 3、已知甲杯中有红黄各一个, 乙杯中有红黄各一个, 现从甲乙两杯各取一球, 在取到红球的条件下, 求取出有红球也有黄球概率?

解: 设事件 A={取到红球}, 事件 B={取红黄球各 1 个}

$$N(A)=3, n(AB)=2, \text{故 } P(B|A) = \frac{2}{3}$$

(三)独立事件的概率

引例: 设 $A_1 = \{\text{投第 1 次骰子, 点数大于 2}\}$, $B = \{\text{投第 2 次骰子, 点数为偶数}\}$

$$P(B|A) = \frac{4 \times 3}{4 \times 6}, P(B) = \frac{3}{6}, P(B|A) = P(B)$$

$$P(AB) = \frac{4 \times 3}{6 \times 6}, P(A) = \frac{4}{6}, P(B) = \frac{3}{6}, P(AB) = P(A)P(B)$$

1、若事件 A 是否发生对事件 B 没有影响, B 是否发生对事件 A 也没有影响, 这样的两个事件叫做相互独立事件。

2、若 A、B 独立

(1) $P(B|A) = P(B)$ (2) $P(AB) = P(A)P(B)$ (3) A、 $\overline{A}$ 之一与 B、 $\overline{B}$ 之一也独立

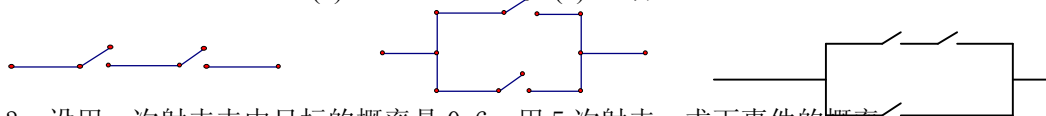
甲、乙二射击运动员分别对一目标射击 1 次, 甲射中的概率为 0.8, 乙射中的概率为 0.9 求下列概率:

(1) 2 人都射中 (2) 2 人中恰有 1 人射中 (3) 2 人至少有 1 人射中

练习、改为甲击中目标的概率是 0.6, 乙击中目标的概率是 0.5

例 2、如图, 每个开关闭合的概率都是 0.7, 计算下列各电路正常工作的概率。

(1) (2) (3) (答 0.847)



例 3、设甲一次射击击中目标的概率是 0.6, 甲 5 次射击, 求下事件的概率

(1) 在第 1 次与第 2 次击中目标, (2) 恰好在 2 次击中目标

例 4: 据报道, 有 10% 的人对某药有胃肠道反应。为考察某厂的产品质量, 现选 5 名患者服用此药, 试求下列事件的概率。

(1) 有人有反应; (2) 不超过 2 人有反应; (3) 至少有 3 人有反应。

例 5、甲、乙两个乒乓球运动员进行乒乓球单打比赛, 已知每一局甲胜的概率为 0.6, 乙胜的概率为 0.4

(1) 求五局三胜制中甲以 3:0, (2) 在五局三胜制中以 3:2 获胜的概率;

(四)全概率公式

设试验 E 的样本空间 Ω 表示为 n 个互斥事件 $B_1, B_2, \dots, B_n$ 的和. $\Omega = B_1 + B_2 + \dots + B_n$.

A 是任意一个事件. 则 $A = AB_1 + AB_2 + \dots + AB_n$, 于是

$$P(A) = P(AB_1) + P(AB_2) + \cdots + P(AB_n) \\ = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n)$$

例 1、设一医院药房中的某种药品是由三个不同的药厂生产的，其中一厂、二厂、三厂生产的药品分别占 $1/4$ 、 $1/4$ 、 $1/2$ 。已知一厂、二厂、三厂生产的药品次品率分别为 7% ， 5% ， 4% 。现从中任取一药品，试求该药品是次品的概率。

解：令 $A = \{\text{该药品是次品}\}$ （显然 A 是一复杂事件）， $B_i = \{\text{药品是由 } i \text{ 厂生产的}\}$ （ $i = 1, 2, 3$ ），显然它们构成一完备事件组，且事件 A 只能与其中之一事件同时发生。故用全概率公式计算。

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) \\ + P(B_3)P(A|B_3) \\ = 0.07 \times 0.25 + 0.05 \times 0.25 + 0.04 \times 0.50 \\ = 0.05$$

例 2：某地成年人体重肥胖者(A_1)占 0.1 ，中等者(A_2)占 0.82 ，瘦小者(A_3)占 0.08 ，又肥胖者、中等者、瘦小者患高血压病的概率分别为 0.2 ， 0.1 ， 0.05 。求该地成年人患高血压的概率。

解：令 $B = \{\text{某人患高血压}\}$ （显然 B 是一复杂事件）， $A_i = \{\text{某人体重的特征}\}$ （ $i = 1, 2, 3$ ），显然它们构成一完备事件组，且事件 B 只能与其中之一事件同时发生。故用全概率公式计算。

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \\ P(B) = 0.1 \times 0.2 + 0.82 \times 0.1 + 0.08 \times 0.05 = 0.106$$

例 3、商店销售一批收音机共 10 台，其中有 3 台次品，但是已经出售了两台。问从剩下的收音中任取一台是正品的概率是多少？

分析：用出售了两台的次品数对样本空间做划分，于样本空间可划分成：出售 0 台次品，1 台次品，2 台次品。分别记为 B_1, B_2, B_3 。设 $A = \{\text{剩下的收音中任取一台是正品}\}$ 。

$$\text{则 } P(B_1) = \frac{C_7^2}{C_7^2} = \frac{1}{15}, P(B_2) = \frac{C_2^1 C_7^1}{C_7^2} = \frac{7}{15}, P(B_3) = \frac{C_2^2 C_7^0}{C_7^2} = \frac{7}{15} \\ P(A|B_1) = \frac{7}{8}, P(A|B_2) = \frac{6}{8}, P(A|B_3) = \frac{5}{8}$$

$$P(A) = P(AB_1) + P(AB_2) + P(AB_3) \\ = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) \\ = 0.7$$

全概率公式其实就是分类乘积概率问题

例 4、盒中有 12 个乒乓球，其中有 9 个是新的。第一次比赛时从其中任取 3 个，赛后放回。第二次比赛时，再从盒中任取 3 个，求第二次取出 3 个球都是新球的概率

解： $A_i = \{\text{第一次取球取到 } i \text{ 个新球}\}$ ， $A = \{\text{第二次取出的 3 个球都是新球}\}$

$$P(A) = P(AB_0) + P(AB_1) + P(AB_2) + P(AB_3) \\ \text{则 } = P(B_0)P(A|B_0) + P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) \\ = 0.9$$

(五)贝叶斯公式

设试验 E 的样本空间 Ω 表示为 n 个互斥事件 $B_1, B_2, \dots, B_n$ 的和. $\Omega = B_1 + B_2 + \dots + B_n$.

A 是任意一个事件. 则

$$P(B_i | A) = \frac{P(B_i A)}{P(A)} = \frac{P(B_i)P(A | B_i)}{P(B_1)P(A | B_1) + P(B_2)P(A | B_2) + \dots + P(B_n)P(A | B_n)}$$

例 1 一个有 5 个选择的考题, 其中只有一个选择正确的. 假定应考人知道正确答案的概率为 p . 如果他最后选对了, 问他确实知道答案的概率是多少?

解: 设 $A = \{\text{知道答案}\}$, $B = \{\text{选则正确}\}$, 由题意可知:

$$P(B | \bar{A}) = \frac{1}{5}, P(B | A) = 1, P(AB) = P(A) = p, \text{由全概率公式:}$$

$$P(B) = P(B | A)P(A) + P(B | \bar{A})P(\bar{A}) = p + \frac{1}{5}(1 - p) = \frac{4p + 1}{5}$$

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{5p}{4p + 1}$$

例 2 某一地区患有癌症的人占 0.005, 患者对一种试验反应是阳性的概率为 0.95, 正常人对这种试验反应是阳性的概率为 0.04, 现抽查了一个人, 试验反应是阳性, 问此人是癌症患者的概率有多大?

求解如下: 设 $C = \{\text{抽查的人患有癌症}\}$,

$$A = \{\text{试验结果是阳性}\}, \text{已知 } P(C) = 0.005, P(\bar{C}) = 0.995,$$

$$P(A | C) = 0.95, \quad P(A | \bar{C}) = 0.04 \text{ 求 } P(C | A).$$

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

$$P(C | A) = 0.1066$$

例 3: 某地成年人体重肥胖者(A_1)占 0.1, 中等者(A_2)占 0.82, 瘦小者(A_3)占 0.08, 又肥胖者、中等者、瘦小者患高血压病的概率分别为 0.2, 0.1, 0.05. 若已知某人患高血压病, 他最可能属于哪种体型。

解: 令 $B = \{\text{某人患高血压}\}$ (显然 B

是一复杂事件), $A_i = \{\text{某人体重的特征}\} (i = 1, 2, 3)$, 显然它们构成一完备事件组, 且事件 B 只能与其中之一事件同时发生. 故用全概率公式计算。

$$P(B) = P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + P(A_3)P(B | A_3) \\ = 0.1 \times 0.2 + 0.82 \times 0.1 + 0.08 \times 0.05 = 0.106$$

$$P(A_1 | B) = \frac{P(A_1)P(B | A_1)}{P(B)} = \frac{0.1 \times 0.2}{0.106} = 0.189$$

$$P(A_2 | B) = \frac{P(A_2)P(B | A_2)}{P(B)} = \frac{0.82 \times 0.1}{0.106} = 0.774$$

$$P(A_3 | B) = \frac{P(A_3)P(B | A_3)}{P(B)} = \frac{0.08 \times 0.05}{0.106} = 0.038$$

第二节 离散型随机变量的分布列

例、设甲一次射击击中目标的概率是 0.6，甲共射击 5 次，设击中目标次数为 X ，求下事件的概率

(1)求 $X=2$ 的概率(2) $P(X=3)$ (2) $P(X=k)$

一、随机变量

1、定义：如果随机试验的结果可划分为若干个互斥事件，每个事件对应于一个实数 X ，则实数 X 就叫做随机变量

2、离散型随机变量：值可以一一列出的随机变量 X 叫离散型随机变量

3、有限离散型随机变量的的概率分布列

$$\begin{array}{c|c} X & x_1 & x_2 & \cdots & x_n \\ \hline P & p_1 & p_2 & \cdots & p_n \end{array}$$

(1) $X=P_i (i=1, 2, \cdots, n)$ 两两互斥；

(2) $p_1 + p_2 + \cdots + p_n = 1$.

例 1、从一批含有 13 只正品，2 只次品，从中抽取 3 只，设抽得次品数为 ξ ，求 ξ 的分布列

解： ξ 的可能的值有 0, 1, 2

$$P(\xi = 0) = \frac{22}{35} \quad P(\xi = 1) = \frac{12}{35} \quad P(\xi = 2) = \frac{1}{35}$$

例 2、随机变量 ξ 的分布列如下：

$P(\xi \geq 2) = 0.7$ ，求 $P(\xi = 1)$

(二)四种重要的分布

1、两点分布：分布列是

叫两点分布

$$\begin{array}{c|c} X & 1 & 0 \\ \hline P & p & 1-p \end{array}$$

2、超几何分布：在 M 件产品中含有 N 件次品，

从中任取 n 件，则次品件数 X 的值有 0, 1, 2, ..., m ，这里 $m = \min\{n, N\}$

其分布列是 $P(X = k) = \frac{C_N^K C_{N-M}^{n-K}}{C_M^n} (k = 0, 1, 2, \cdots, m)$ 这种分布叫做超几何分布

3、二项分布：在一次试验中事件 A 发生的概率为 $P(A) = p$ ，则 n 次独立重复中试验事件 A 发生次数 X 的概率分布叫做二项分布，记作 $X \sim B(n, p)$ ，记 $1 - p = q$

其分布是

$$\begin{array}{c|c} \xi & 0 & 1 & 2 & \cdots & n \\ \hline P & q^n & C_n^1 p^1 q^{n-1} & C_n^2 p^2 q^{n-2} & \cdots & p^n \end{array}$$

4、在一次试验中事件 A 发生的概率为 $P(A) = p$ ，事件 A 第一次发生试验的次数 X 的概率分布叫几何分布. 记 $1 - p = q$ ，其分布列是

$$\begin{array}{c|c} \xi & 1 & 2 & 3 & \cdots & n & \cdots \\ \hline P & p & qp & q^2 p & \cdots & q^{n-1} p & \cdots \end{array}$$

例 1、有一批种子，每粒发芽的概率为 $\frac{2}{3}$ ，播下 5 粒种子，计算：

(I) 求恰好有 ξ 粒发芽的分布列；(II) 至少有 4 粒发芽的概率；

(III) 至少有 1 粒发芽的概率.

例 2、一盒子内有红球 3 个，白球 5 个，从中摸 2 个球，若至少摸到一个红球，则中奖，求中奖的概率

解：设摸到红球个数为 X ，则 X 服从超几何分布

$$P(X \geq 1) = \frac{C_3^1 C_5^1}{C_8^2} + \frac{C_3^2}{C_8^2} = \frac{9}{14}$$

解2：中奖的概率为 $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{C_5^2}{C_8^2} = \frac{9}{14}$

第三节、期望与方差

一、定义式

1、期望(均值)： $EX = x_1P_1 + x_2P_2 + \dots + x_nP_n$

2、方差： $DX = (x_1 - EX)^2 p_1 + (x_2 - EX)^2 p_2 + \dots + (x_n - EX)^2 p_n$

例1. 随机变量 η 分布列如下：若 $E\eta = 2$, 求 $D\eta$

η	1	2	3
p	a	b	0.3

例2、一盒子内装有7个乒乓球，其中3个旧的，4个新的
从中任意取4个，取到旧球的个数为 X ，求 $E(X)$ ， $D(X)$

例3、张三用10万元投资一个项目，成功可获利3万元，
失败将损失1万元，成功的概率80%，失败的概率20%
求此投资获利平均值

例4. 某会馆搞抽奖活动，每人交30元的入场费. 抽奖规则是：从一个装有分值分别为1,2,3,4,5,6的六个小球的抽奖箱中，有放回的抽取两次，每次抽取一个球。若抽得两球的分值之和为12分则获一等奖品；若抽得两球的分值之和为11分或10分则获得二等奖品。若抽得两球的分值之和低于10分，则不获奖。

(1) 求每位会员获奖的概率；(2) 假设二等奖品的价值为100元，会馆不赔不赚，求一等奖品的价值

解：(1) 解：设 $A = \{\text{会员获奖}\}$ ，则 $P(A) = \frac{1+2+3}{6 \times 6} = \frac{1}{6}$ ，所以每位会员获奖的概率为 $\frac{1}{6}$ 。

(2) 设一等奖品的价值 m 元，每位来宾获奖价值 ξ 元，则 ξ 的值有： m 、100、0

$$P(\xi = m) = \frac{1}{36}, \quad P(\xi = 100) = \frac{2+3}{36} = \frac{5}{36}, \quad P(\xi = 0) = 1 - \frac{1}{36} - \frac{5}{36} = \frac{5}{6}$$

$$E\xi = \frac{1}{36} \times m + \frac{5}{36} \times 100 + \frac{5}{6} \times 0 = \frac{m+500}{36}, \quad \text{令 } E\xi = \frac{m+500}{36} = 30, \text{ 得 } m=580$$

答：(1) 每位会员获奖的概率为 $\frac{1}{6}$ 。(2) m 应为 580 元。

二、公式

1、设 x 服从两点分布, 则 $E(X) = P$, $Dx = P(1-P)$

2、设 $X \sim B(n, p)$, 则 $E\xi = nP$, $D\xi = nP(1-P)$

3、 $DX = E(X - EX)^2 = EX^2 - (EX)^2$

2的证明：设 $q = 1 - p$

$$\begin{array}{ccccccc} \xi & 0 & 1 & 2 & \dots & n \\ P & q^n & C_n^1 p^1 q^{n-1} & C_n^2 p^2 q^{n-2} & \dots & p^n \end{array}$$

$$\begin{aligned} EX &= 0 \times q^n + 1 \times C_n^1 p^1 q^{n-1} + 2 \times C_n^2 p^2 q^{n-2} + \dots + n C_n^n p^n \\ &= np(C_{n-1}^0 q^{n-1} + C_{n-1}^1 p^1 q^{n-2} + \dots + C_{n-1}^{n-1} p^{n-1}) = np(p+q)^{n-1} = np \end{aligned}$$

$$k C_n^k p^k q^{n-k} = np C_{n-1}^{k-1} p^{k-1} q^{n-k}$$

$$C_n^0 q^n + C_n^1 p^1 q^{n-1} + C_n^2 p^2 q^{n-2} + \dots + C_n^n p^n = 1$$

$$1 \times C_n^1 p^1 q^{n-1} + 2 \times C_n^2 p^2 q^{n-2} + \dots + n C_n^n p^n = np$$

$$\begin{aligned} EX^2 &= 0^2 \times q^n + 1^2 \times C_n^1 p^1 q^{n-1} + 2^2 \times C_n^2 p^2 q^{n-2} + \dots + n^2 C_n^n p^n \\ &= np(1 \times C_{n-1}^0 q^{n-1} + 2 \times C_{n-1}^1 p^1 q^{n-2} + \dots + n C_{n-1}^{n-1} p^{n-1}) \end{aligned}$$

$$= np[(n-1)p + 1] = np(np - p + 1)$$

$$\begin{aligned} DX &= EX^2 - (EX)^2 = np(np - p + 1) - (np)^2 \\ &= np(np - p + 1 - np) = np(1 - p) \end{aligned}$$

η	1	2	3
p	a	0.5	0.3

例 1. 随机变量 η 分布列如下: $\xi = 4\eta + 3$, 求 $E\xi, D\xi$

解: $a + 0.5 + 0.3 = 1, a = 0.2$,

$$E\eta = 0.2 + 1 + 0.9 = 2.1, D\eta = E\eta^2 - (E\eta)^2 = 0.2 + 2 + 2.7 - 2.1^2 = 0.49$$

$$E\xi = E(4\eta + 3) = 4E\eta + 3 = 4 \times 2.1 + 3 = 11.4$$

$$D\xi = D(4\eta + 3) = 4^2 D\eta = 16 \times 0.49 = 7.84$$

例 2. 一袋内有 3 个红球和 4 个白球, 每次一球取后放回, 取 5 次取到红球的个数为 X , 求 EX, DX

解: 由于 $X \sim B(5, \frac{3}{7})$, 于是 $EX = np = \frac{15}{7}$, $DX = npq = 5 \times \frac{3}{7} \times \frac{4}{7} = \frac{60}{49}$

例 3. 某种种子每粒发芽的概率都为 0.9, 现播种了 1000 粒, 对于没有发芽的种子, 每粒需再补种 2 粒, 补种的种子数记为 X , 求 EX, DX

解: 没有发芽的种子数为 ξ , 则 $\xi \sim B(1000, 0.1)$, $E\xi = np = 100$, $D\xi = npq = 90$,

因 $X = 2\xi$, 故 $EX = E(2\xi) = 2E\xi = 2 \times 100 = 200$, $DX = 4D\xi = 4 \times 90 = 360$

例 4. 甲乙玩“石头、剪刀、布”, 双方出三种手势是等可能的

(1) 求出在 1 次游戏中甲胜乙的概率;

(2) 若甲、乙双方共进行了 3 次游戏, 其中甲胜乙的次数记作 X , 求 $E(X), D(X)$

解: (1) 设 A 在 1 次游戏中甲胜乙的概率 $P = \frac{3}{9} = \frac{1}{3}$

(2) $X \sim B(3, \frac{1}{3})$, 于是 $E\xi = np = 1$, $D\xi = npq = \frac{2}{3}$

例 5. 周末 3 人都从甲, 乙, 丙三地的选择一个地方游玩, 每个人选甲, 乙, 丙的概率都分别为 $\frac{1}{6}, \frac{1}{3}, \frac{1}{2}$, 用 ξ 表示选择在乙地游玩的人数

(1) 求 $E\xi, D\xi$; (2) 求三地都有人选择的概率

解: 由于 $\xi \sim B(4, \frac{1}{3})$, 于是 $E\xi = nP = \frac{4}{3}$, $D\xi = nP(1-P) = 4 \times \frac{1}{3} \times \frac{2}{3} = \frac{8}{9}$

例 6. 在备选的 10 道题中, 甲能答对其中的 6 道题, 规定从备选题中随机抽出 3 道题进行测试, 至少答对 2 道题则及格. 答对 3 题则优秀 (1) 求甲及格的概率; (2) 已知甲及格, 求甲优秀的概率 (3) 设甲答对试题数 ξ 的数学期望. (4) 在备选的 10 道题中, 乙能答对其中的 8 道题, 求甲乙及格的概率

解: (1) 设 $A = \{\text{甲测试的及格概率}\}$, 则 $P(A) = \frac{C_6^2 C_4^1 + C_6^3}{C_{10}^3} = \frac{80}{120} = \frac{2}{3}$

(2) ξ 的值有: 0, 1, 2, 3, 则 $P(\xi = 0) = \frac{C_4^3}{C_{10}^3} = \frac{4}{120} = \frac{1}{30}$, $P(\xi = 1) = \frac{C_6^1 C_4^2}{C_{10}^3} = \frac{36}{120} = \frac{3}{10}$

$P(\xi = 2) = \frac{C_6^2 C_4^1}{C_{10}^3} = \frac{60}{120} = \frac{1}{2}$, $P(\xi = 3) = \frac{C_6^3}{C_{10}^3} = \frac{20}{120} = \frac{1}{6}$

$$E\xi = 0 \times \frac{1}{30} + 1 \times \frac{3}{10} + 2 \times \frac{1}{2} + 3 \times \frac{1}{6} = 1.8$$

例 7. 有特殊骰子: 1 面 0 点, 2 面 1 点, 三面 2 点. 投 2 次骰子, 求点数之积的分布列

三、泊松分布

1、泊松定理

设随机变量 $X_n \sim B(n, p_n)$, $(n = 1, 2, 3, \dots)$, $E(X_n)$ 为常数 λ , 即 $np_n = \lambda$.

$$\text{则 } \lim_{n \rightarrow \infty} P\{X_n = k\} = \frac{\lambda^k e^{-\lambda}}{k!}$$

证: 由 $np_n = \lambda$ 得 $p_n = \frac{\lambda}{n}$

$$\begin{aligned}\lim_{n \rightarrow \infty} P\{X_n = k\} &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{\lambda^k}{k!} \left[1 \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right] \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \times 1 \times e^{-\lambda} \times 1 = \frac{\lambda^k e^{-\lambda}}{k!}\end{aligned}$$

由此可见, 在 $E(X_n) = \text{常数 } \lambda$ 的前提下, 当 n 很大时, $\lim_{n \rightarrow \infty} P\{X_n = k\} = \frac{\lambda^k e^{-\lambda}}{k!}$

2、泊松分布

如果随机变量 X 的概率分布为 $P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!} (k = 0, 1, 2, \dots), \lambda > 0$, 则称 X 服从泊松分布, 记作 $X \sim P(\lambda)$

3. 设 X 服从参数为 $\lambda (\lambda > 0)$ 的泊松分布, 即 $X \sim P(\lambda)$, X 的分布律为

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots \quad \text{则} \quad E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

$$\stackrel{k-1=l}{=} \lambda e^{-\lambda} \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \quad E(X^2) = E[X(X-1) + X] = E[X(X-1)] + \lambda,$$

$$\text{而} \quad E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}$$

$$\stackrel{k-2=l}{=} \lambda^2 e^{-\lambda} \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2,$$

$$\text{从而} E(X^2) = \lambda^2 + \lambda, \quad D(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

在泊松分布 $P(\lambda)$ 中, 它的唯一参数 λ 既是期望又是方差。

四、常见的几个母函数:

对于离散型随机变量, 常用到他们的母函数。定义: $P_X(t) = \sum t^k P(X = k) = E(t^X)$

$$1、\text{二项分布: } P(t) = \sum_{k=0}^{\infty} C_n^k p^k q^{n-k} t^k = (q + pt)^n$$

$$2 \text{ 几何分布: } P(t) = \sum_{k=1}^{\infty} q^{k-1} p t^k = pt \sum_{k=1}^{\infty} (qt)^{k-1} = \frac{pt}{1-qt}$$

$$3. \text{泊松分布: } P(t) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} t^k = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$$

第四节 统计初步

复习、总体与样本

(1) 总体与个体; (2) 样本与样本容量;

(3) 统计学的基本思想方法: 用样本的情况去估计总体的相应情况

例 1. 某次考试有 70000 名学生参加, 为了了解这 70000 名考生的数学成绩, 从中抽取 1000

名考生的数学成绩进行统计分析，在这个问题中，

(1)总体是_____总体容量是_____

(2)样本是_____样本容量是_____

一、抽样方法：

(一)简单随机抽样

特点：①总体容量有限②逐个抽取不放回③每个个体被抽到的机会相等

1、抽签法：编号，做签，抽签

2、随机数表法：编号(每个号的数码数相同)，在机数表中任选一数头号

例 1. 下列抽样的方式是否属于简单随机抽样？为什么？

(1) 从无限多个个体中抽取 50 个个体作为样本。

(2) 箱子里共有 100 个零件，从中选出 10 个零件进行质量检验，在抽样操作中，从中任意取出一个零件进行质量检验后，再把它放回箱子。

例 2. 某车间工人加工一种轴 100 件，为了了解这种轴的直径，要从中抽取 10 件轴在同一条件下测量，如何采用简单随机抽样的方法抽取样本？

解法 1: (抽签法) 将 100 件轴编号为 1, 2, ..., 100, 并做好大小、形状相同的号签，分别写上这 100 个数，将这些号签放在一起，进行均匀搅拌，接着连续抽取 10 个号签，然后测量这个 10 个号签对应的轴的直径。

解法 2: (随机数表法) 将 100 件轴编号为 00, 01, ..., 99, 在随机数表中选定一个起始位置，如取第 21 行第 1 个数开始，选取 10 个为 68, 34, 30, 13, 70, 55, 74, 77, 40, 44, 这 10 件即为所要抽取的样本。

(二)系统抽样：编号，分段(取 k 个分 k 段), 选头号加段距选下一个

例 1. 从 624 人中用系统抽样的方法抽出 10%, 怎样抽？

解：第一步：将 624 人编号(000,001,..., 623),用随机数表法抽 4 个号码剔除 4 人

第二步：将剩下的 620 人重新编号 (1, 2, 3, ..., 620), 并分成 62 段；

第三步：在第一段 1, 2, ..., 9, 10 中用抽签法抽出号码 a；

则编号为 a, a+10, a+20, ..., a+610 的个体组成样本。

(三)分层抽样：按比例抽样

$$a:b:c = A:B:C, \quad \frac{a}{A} = \frac{b}{B} = \frac{c}{C} = \frac{n}{N}$$

例 1、高中学生有 900 人，其中高一 300 人，高二 200 人，高三 400 人，现采用分层抽取容量为 45 人的样本，那么高一、高二、高三年级抽到的人数分别为_____15, 10, 20

例 2、有老中青 430 人,其中青年 160 人，中年人数是老人的 2 倍。现采用分层抽样方法进行调查，抽到青年 32 人，则抽到老年人数____人 18

例 3、A、B、C 三种型号数量依次为 1600, 1600, 4800. 用分层抽样投到 A 种型 16 件，则样本的容量_____件. 80

问题一：逐个抽取不放回每个个体被抽到的机会相等？例如从 5 个选两个，则每个被选到的概率都是 2/5。要注意学生误解为第一个选到某号的条件下选第二个号的概率

二、频率分布直方图

(一)频率分布直方图

例 1、一样本数据如下(分):

61,76,70,56,81,91,92,91,75,81,88,67,101,103,95,91, 77,86,81,83,82,82,64,79,86,85,75,71,49,45,
作出频率分布直方图

解：(1)分组(2)列出频率分布表(3)作出频率分布直方图

1、极差=最大数-最小数，组数，组距

2、横轴：表示样本数值 3、纵轴：频率/组距

3、直方图的高=频率/组距，直方图的面积和=1

4、中位数≈平分直方图面积的横坐标

练习、一个容量为 20 的样本数据，数据的分组及各组的频数如下：

[10, 20), 2; [20, 30), 3; [30, 40), 4; [40, 50), 5;

[50, 60), 4; [60, 70], 2;

则样本在区间 $(-\infty, 50)$ 上的频率为 ()

A. 0.5

B. 0.7

C. 0.25

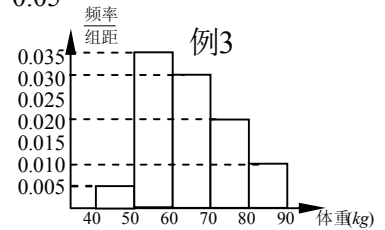
D. 0.05

例 2、一样本频率分布直方图(如图)。

(1)若体重[40,60)的有 40 人, 则[70,90]的同学有_____人

(2)若要从体重在[60,70), [70,80), [80,90]三组内, 用分层抽样的方法取 12 人参加某项活动, 则体重在[60,80)内的要抽取_____人

(3) 样本中位数 $\approx$ _____



(二) 频率分布折线图与总体密度曲线

1、频率分布折线图: 连接频率分布直方图中各小矩形上端中点的折线

2、总体密度曲线: 当总体容量很大时, 如果样本容量越来越大趋于总体容量, 组距越来越小趋于 0, 则相应的频率折线图将趋于一条光滑的曲线, 这条光滑的曲线叫做总体的密度曲线.

(三) 频数分布茎叶图

例 1. 甲、乙两篮球运动员每场比赛的得分如下:

甲 13,51,23,8,26,38,16,33,14,28,39.

乙 49,24,12,31,31,50,31,44,36,15,37,25,36,39

试比较这两位运动员的得分水平.

解: 画茎叶图

由茎叶图知

甲的分数集中于 20 多分

乙的分数集中于 30 多分

于是可知乙得分水平更好

甲		乙
8	0	
463	1	25
368	2	54
389	3	161679
	4	49
1	5	0

三、样本重要数量

1、众数: 样本数据中出现次数最多的数

2、中位数: 样本数据按从小到大的顺序排列中处于中间位置的一个数, 或两个中间位置的数的平均数.

3、平均数 $\bar{x} = \frac{1}{n}(x_1 + x_2 + x_3 + \cdots + x_n)$

4、方差 $s^2 = \frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2]$, 标准差 s 。注: 方差大说明波动大。

例 1、五个数 1, 1, 5, 6, a 的平均数是 3, 则 $a =$ _____, 众数=_____, 中位数=_____, 标准差=_____.

例 2. 判定正误

①若样本众数是 5, 则样本中出现次数最多的数是 5;

②若样本平均数是 0, 则样本的中位数为 0;

③若样本 1, 2, x , 4 的中位数是 3, 则 $x = 4$;

④若样本的平均数是正数, 则样本的每个数据都是正数

例 3. 如果 5 个数 x_1, x_2, x_3, x_4, x_5 的平均数是 7 方差是 10, 那么 $2x_1+1, 2x_2+1, 2x_3+1,$

$2x_4+1, 2x_5+1$ 这 5 个数的平均数=_____15, 方差=_____40

例 4、对两运动员甲、乙二人在相同的条件下进行了 6 次测试, 测得他们速度如下:

甲: 27, 38, 30, 37, 35, 31; 乙: 33, 29, 38, 34, 28, 36

根据以上数据, 试判断他们谁更优秀。

解: $\bar{x}_甲 = 33, \bar{x}_乙 = 33, s_甲^2 = 18.8, s_乙^2 = 15.2,$

茎	叶
8	9 7
9	3 1 6 4 0 2

因为 $\overline{x_{\text{甲}}} = \overline{x_{\text{乙}}}$, $S_{\text{甲}}^2 > S_{\text{乙}}^2$ 所以乙比甲更稳定, 乙比甲更优秀。

例 5、8 个班参加合唱比赛的得分茎叶图,

则这组数据的中位数和平均数分别是()

A、91.5 和 91.5 B、91.5 和 92 C、91 和 91.5 D、92 和 92

例 6. 一样本频率分布直方图(如图)

(1) 众数 $\approx$ ____ (2) 平均体重数 $\approx$ ____

*众数 $\approx$ 最高直方图的小组横坐标中点

平均数 $\approx$ 直方图面积 $\times$ 底边中点横坐标的总和

四、正态分布

(一) 正态分布的认识

1、频率分布折线图: 连接频率分布直方图中各小矩形上端中点的折线

2、总体密度曲线: 当样本容量 $n \rightarrow \infty$, 组距 $d \rightarrow 0$ 时, 则频率分布折线图将趋于一条光滑的曲线, 这条光滑的曲线叫做总体的密度曲线.

3、设实数集 R 上连续型随机变量 X 的密度曲线有中间高, 两边低, **轴对称的特征**, 则随机变量 X 服从正态分布, 记为 $X \sim N(\mu, \sigma^2)$, 这里 $\mu = E(X)$, $\sigma = D(X)$

4、正态分布的在区间中的概率:

$$(1) P(X \in R) = 1, P(X < \mu) = P(X \leq \mu) = P(X > \mu) = P(X \geq \mu) = \frac{1}{2}$$

$$(2) P(a \leq X \leq b) = \text{阴影面积}, P(\mu - \sigma < X \leq \mu + \sigma) = 0.6826,$$

$$P(\mu - 2\sigma < X \leq \mu + 2\sigma) = 0.9544 \quad P(\mu - 3\sigma < X \leq \mu + 3\sigma) = 0.9974$$

$$5、\text{密度函数 } f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in (-\infty, +\infty)$$

例 1、随机变量 $X \sim N(0, 1)$, $P(\xi > 2.5) = a$, 求 $P(-2.5 < \xi < 0)$

例 2. 设随机变量 ξ 服从正态分布 $N(3, 4)$, 若 $P(\xi < 2a - 3) = P(\xi > a + 2)$, 求 a 的值

例 3、随机变量 $\xi \sim N(3, 100)$, 求 $P(\xi < -7)$

例 4、某次地区 20000 万名学生的考试的成绩 $\xi \sim N(90, 49)$, 单位为分, 求成绩在 (90, 111] 之间的人数

例 5. 某年级 1000 名学生的一次信息技术成绩近似服从于正态分布 $N(70, 100)$, 不低于 80 分为优秀, 求优秀的学生人数

例 6、有生产线生产的铁钉的长度 $\xi \sim N(10, 0.04)$ (单位: cm), 在某时间段取一枚测得长度为 9.3 cm, 问此时间段生产是否正常?

例 7、 $X \sim N(\mu, \sigma^2)$, 其密度函数为 $f(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{x^2}{8}}$, 则 $\mu = \underline{0}$, $\sigma = \underline{2}$

(4) 在直线 $x=m$ 左侧的面积就是 $P(x \leq m)$ 的值, 记作 $F(m)$, 即 $F(m) = P(x \leq m) =$ 左侧面积

注意: 对于连续型总体或随机变量 $P(x \leq m) = P(x < m)$

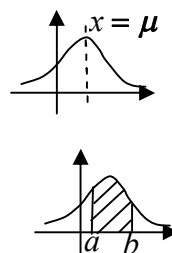
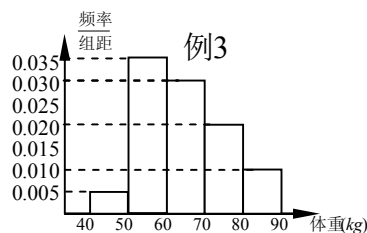
(二) 标准正态分布

1、当 $\mu = 0$, $\sigma = 1$ 时的正态分布, 叫做标准正态分布记作 $N(0, 1)$, 标准正态分

布密度函数是 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in R$

2、在标准正态分布 $N(0, 1)$ 下, $F(m) = P(x \leq m) = \int_{-\infty}^m \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, 可记作 $\Phi(m)$,

3、 $\Phi(m)$ 的求法



①当 $m > 0$ 时, $\Phi(m)$ 可以由《标准正态分布表》查到

②当 $m < 0$ 时, $p(x < m) = p(x > -m) = 1 - p(x < -m)$, 即 $\Phi(m) = 1 - \Phi(-m)$

③ $p(x_1 < x < x_2) = \Phi(x_2) - \Phi(x_1)$

例 1、求标准正态总体在 $(-1, 2)$ 内取值的概率。

解: 利用公式 $p(x_1 < x < x_2) = \Phi(x_2) - \Phi(x_1)$ 有

$$p(-1 < x < 2) = \Phi(2) - \Phi(-1) = \Phi(2) - \{1 - \Phi[-(-1)]\}$$

$$4、\text{设 } X \sim N(\mu, \sigma^2), F(m) = P(x \leq m) = \int_{-\infty}^m \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$\text{则 } F(m) = \Phi\left(\frac{m-\mu}{\sigma}\right), P(x_1 < x < x_2) = F(x_2) - F(x_1) = \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right)$$

证明: 在 $F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$ 中,

设 $y = \frac{x-\mu}{\sigma}$, 则 $dx = \sigma dy$, 积分上限 $= \frac{m-\mu}{\sigma}$

$$F(x) = \int_{-\infty}^{\frac{m-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2}} \sigma dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{m-\mu}{\sigma}} e^{-\frac{y^2}{2}} dy = \Phi\left(\frac{m-\mu}{\sigma}\right)$$

$$\text{于 } P(x_1 < x < x_2) = F(x_2) - F(x_1) = \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right)$$

例 1、已知正态总体 $N(1, 4)$, 求 $P(x < 3)$ 。

$$\text{解: } P(x < 3) = F(3) = \Phi\left(\frac{3-1}{2}\right) = \Phi(1) = 0.8413。$$

例 2 分别求正态总体 $N(\mu, \sigma^2)$ 在区间 $(\mu - \sigma, \mu + \sigma)$ 、 $(\mu - 2\sigma, \mu + 2\sigma)$ 、 $(\mu - 3\sigma, \mu + 3\sigma)$ 内取值的概率。

$$\text{解: } P(\mu - \sigma < x < \mu + \sigma) = F(\mu + \sigma) - F(\mu - \sigma) = \Phi\left(\frac{(\mu + \sigma) - \mu}{\sigma}\right) - \Phi\left(\frac{(\mu - \sigma) - \mu}{\sigma}\right)$$

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 2 \times 0.8413 - 1 \approx 0.683;$$

$$P(\mu - 2\sigma < x < \mu + 2\sigma) = F(\mu + 2\sigma) - F(\mu - 2\sigma) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.954;$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = F(\mu + 3\sigma) - F(\mu - 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 \approx 0.997$$

第二章 随机变量的一般理论

第一节 随机变量的分布函数与密度函数

一、随机变量的分布函数

1、设 X 是一个随机变量, x 是任一实数, 则称函数 $F(x) = P(X \leq x), x \in (-\infty, +\infty)$ 为随机变量 X 的分布函数。

例 1、设随机变量 X 的分布列如表

(1) 求 X 的分布函数 $F(x)$

(2) 求 $P(X > 2), P(0.5 < X \leq 6)$

X	1	2	3
p	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{2}$

$$\text{解(1) } P(X < 1) = 0, P(X \leq x, 1 \leq x < 2) = P(X = 1) = \frac{1}{5}$$

$$P(X \leq x, 2 \leq x < 3) = P(X = 1) + P(X = 2) = \frac{1}{5} + \frac{3}{10} = \frac{1}{2}$$

$$P(X \leq x, x \geq 3) = P(X=1) + P(X=2) + P(X=3) = \frac{1}{5} + \frac{3}{10} + \frac{1}{2} = 1$$

$$\text{故 } F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{5}, & 1 \leq x < 2 \\ \frac{1}{2}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

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

$$P(0.5 < X \leq 6) = F(6) - F\left(\frac{3}{2}\right) = 1 - \frac{1}{5} = \frac{4}{5}$$

2、对于任意的实数 $x_1, x_2, P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

例 2、一个靶子是一个半径为 2 的圆盘. 设击中靶上任一同心圆盘的概率与该圆盘的面积成正比, 并设射击都能中靶. 以 X 表示弹着点到圆心的距离. 试求随机变量 X 的分布函数

解当 $x < 0$ 时, $F(x) = P(X \leq x) = 0$

当 $0 \leq x \leq 2$ 时, $F(x) = P(X \leq x) = kx^2$, 故 $F(2) = 4k = 1, k = \frac{1}{4}$, 故 $F(x) = \frac{1}{4}x^2$

$$\text{当 } x \geq 2 \text{ 时 } F(x) = P(X \leq x) = 1, \text{ 故 } F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x^2, & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

二、随机变量的密度函数

1、定义: 若对随机变量 X 的分布函数 $F(x)$, 存在非负函数 $f(x)$, 使对任意实数 x 有

$F(x) = \int_{-\infty}^x f(t)dt$, 则称 X 为连续型随机变量.

2、密度函数 $f(x)$ 的性质

$$(1) f(x) \geq 0 \quad (2) \int_{-\infty}^{+\infty} f(x)dx = 1 \quad (3) P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$$

(4) 若 $f(x)$ 在点 x 处连续, 则有 $f(x) = F'(x)$

$$\text{例 1、随机变量 } X \text{ 的概率密度为 } f(x) = \begin{cases} a \cos x & |x| \leq \frac{\pi}{2} \\ 0 & \text{其它} \end{cases} \text{ 求 } P(0 \leq X \leq \frac{\pi}{4})$$

解 利用密度函数的性质求出 a

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a \cos x dx = 1 \Rightarrow a = \frac{1}{2}$$

$$\text{于是 } P(0 \leq X \leq \frac{\pi}{4}) = \int_0^{\frac{\pi}{4}} \frac{1}{2} \cos x dx = \frac{\sqrt{2}}{4}$$

$$\text{例 2 设随机变量 } X \text{ 具有概率密度 } f(x) = \begin{cases} \frac{1}{2} \cos x & |x| < \frac{\pi}{2} \\ 0 & \text{其他} \end{cases} \text{ 求 } X \text{ 的分布函数}$$

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

$$\text{当 } x < -\frac{\pi}{2} \text{ 时 } F(x) = 0$$

$$\text{当 } -\frac{\pi}{2} \leq x < \frac{\pi}{2} \text{ 时 } F(x) = \int_{-\frac{\pi}{2}}^x \frac{1}{2} \cos x dx = \frac{1}{2}(1 + \sin x)$$

$$\text{当 } x \geq \frac{\pi}{2} \text{ 时 } F(x) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2} \cos x dx = 1$$

于是X的分布函数为

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

例 3、随机变量 X 的分布函数为 $F(x) = \begin{cases} 0 & x \leq 0 \\ x^2 & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$, 求 X 的密度函数

$$\text{解 X 的密度函数为 } f(x) = F'(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{其它} \end{cases}$$

3、均匀分布与指数分布

(1) 若连续型随机变量 X 的概率密度是 $f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其它} \end{cases}$, 则称 X 在区间(a,b)上服从

$$\text{均匀分布. 其分布函数为 } F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

(2) 若连续型随机变量 X 的概率密度是 $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$, 则称 X 服从指数分布.

$$\text{其分布函数为 } F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

三、连续型随机变量的数学期望与方差

1、若 X 是连续型随机变量, 则 $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$, $D(X) = E\{[X - E(X)]^2\}$,

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

2、几个简单公式

(1) 设 C 是常数, 则 $E(C) = C$, $D(X) = 0$ (2) 若 k 是常数, 则 $E(kX) = kE(X)$; $D(X) = k^2 D(X)$

推论若 $Y = aX + b$ 则 $E(Y) = aE(X) + b$, $D(Y) = a^2 D(X)$,

例 1 设 $X \sim U(a, b)$, 求 $E(X)$, $D(X)$

解 X 的概率密度为

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

X 的数学期望为

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

即数学期望位于区间 (a,b) 的中点.

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{1}{3}(b^2 + ab + a^2),$$

$$\text{于是 } D(X) = E(X^2) - [E(X)]^2 = \frac{1}{12}(b-a)^2.$$

例 2 设随机变量 X 服从参数为 $\lambda (\lambda > 0)$ 的指数分布, 其概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \text{ 求 } E(X), D(X)$$

解:

$$\begin{aligned} E(X) &= \int_0^{+\infty} \lambda x e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{+\infty} = \frac{1}{\lambda}. \end{aligned}$$

$$E(X^2) = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = -x^2 e^{-\lambda x} \Big|_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx = \frac{2}{\lambda^2} \text{ (由 } E(X) \text{ 的计算),}$$

$$\text{从而 } D(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

例 3 设连续型随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2), & 0 \leq x \leq 1, \\ 0, & \text{其他,} \end{cases} \text{ 求方差 } D(X).$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x \frac{3}{2}(1-x^2)dx = \frac{3}{8}, \\ E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^1 x^2 \frac{3}{2}(1-x^2)dx = \frac{1}{5}, \end{aligned}$$

从而

$$D(X) = E(X^2) - [E(X)]^2 = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320}.$$

例 5 设随机变量 $X \sim N(\mu, \sigma^2)$, 求方差 $D(X)$ 和标准差 $\sigma(X)$.

解

$$\begin{aligned} D(X) &= E\{X - E(X)\}^2 = E[(X - \mu)^2] \\ &= \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \end{aligned}$$

作变换 $\frac{x-\mu}{\sigma} = t$, 则 $dx = \sigma dt$, $D(X) = \int_{-\infty}^{+\infty} \sigma^2 t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left[-te^{-\frac{t^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt \right] = \sigma^2, \sigma(X) = \sqrt{D(X)} = \sigma.$$

四、随机变量函数的分布

1、设 $f(x)$ 是一个函数, X 是随机变量, 则 $f(X)$ 也是一个随机变量, 记作 $Y = f(X)$, Y 叫做随机变量函数

2、离散型随机变量的分布列

$$\begin{array}{c|c} X & x_1 \quad x_2 \cdots x_n \cdots \\ \hline P & p_1 \quad p_2 \cdots p_n \cdots \end{array}$$

例 1、设 $Y = 2X - 1$, X 的分布列如表 1

则 Y 的分布列如表 2

$$\begin{array}{c|c} Y & 2x_1 - 1 \quad 2x_2 - 1 \cdots 2x_n - 1 \cdots \\ \hline P & p_1 \quad p_2 \quad \cdots p_n \quad \cdots \end{array}$$

例 2、已知 X 概率分布为

X	0	1	2	3	4	5
P	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

求 $Y = (X - 2)^2$ 的概率分布

解: Y 的对应值分布

Y	$(-2)^2$	-1^2	(0^2)	1^2	2^2	3^2
P	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

注意到 $(-2)^2 = 2^2 = 4, (-1)^2 = 1^2 = 1$

于是 $P(Y = 4) = \frac{1}{12} + \frac{2}{9} = \frac{11}{36}, P(Y = 1) = \frac{1}{6} + \frac{1}{12} = \frac{3}{12} = \frac{1}{4}$

则合并化简后得 Y 的分布列如表 2

Y	0	1	4	9
P	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{2}{9}$

3、连续型随机变量的密度函数

例 1、设 $Y = 2X - 1, X \sim N(0, 1)$, 求 Y 密度函数 $f_Y(y)$

解: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$$F_Y(y) = P(Y \leq y) = P(2X - 1 \leq y) = P(X \leq \frac{y+1}{2}) = F_X(\frac{y+1}{2})$$

$$\text{两边对 } y \text{ 取导数得 } f_Y(y) = f_X(\frac{y+1}{2}) (\frac{y+1}{2})' = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(\frac{y+1}{2})^2}{2}} = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(y+1)^2}{8}}$$

这表明 $Y \sim N(-1, 4)$

例 2、设 $Y = X^2, X \sim N(0, 1)$, 求 Y 密度函数 $f_Y(y)$

解: 当 $y < 0$ 时, $X^2 < y$ 是不可能事件于是 $f_Y(y) = 0$

$$\text{当 } y \geq 0 \text{ 时, } F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-\frac{x^2}{2}} dx$$

$$\text{于是 } f_Y(y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{y}{2}} (\sqrt{y})' = \frac{2}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}$$

$$5、E(X+Y)=E(X)+E(Y), E\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n E(X_k).$$

例 1、求二项分布的数学期望

解: $X \sim B(n, p)$, 则 X 表示 n 重贝努里试验中的“成功”次数

$$\text{设 } X_i = \begin{cases} 1 & \text{如第 } i \text{ 次试验成功} \\ 0 & \text{如第 } i \text{ 次试验失败} \end{cases} \quad i=1, 2, \dots, n$$

则 $X = X_1 + X_2 + \dots + X_n$ 因为 $P(X_i = 1) = p$

$$P(X_i = 0) = 1 - p \quad E(X_i) = 1 \cdot p + 0 \cdot (1 - p)$$

$$E(X) = \sum_{i=1}^n E(X_i) = np$$

例 2、把数字 $1, 2, \dots, n$ 任意地排成一列, 如果数字 k 恰好出现在第 k 个位置上, 则称为一个巧合, 求巧合个数的数学期望.

解: 设巧合个数为 X ,

$$\text{设 } X_k = \begin{cases} 1, & \text{数字 } k \text{ 恰好出现在第 } k \text{ 个位置上} \\ 0, & \text{否则} \end{cases} \quad k=1, 2, \dots, n$$

$$\text{则 } X = \sum_{k=1}^n X_k \quad \text{由于} \quad E(X_k) = P(X_k = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

$$\text{于是 } E(X) = \sum_{k=1}^n E(X_k) = n \cdot \frac{1}{n} = 1$$

6、设随机变量 X 与 Y 相互独立, 则有

$$(1) D(X+Y) = D(X) + D(Y). \quad (2) D(aX+bY) = a^2 D(X) + b^2 D(Y), \quad D(X-Y) = D(X) + D(Y).$$

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

$$= E\{[X - E(X)]^2\} + E\{[Y - E(Y)]^2\} + 2E\{[X - E(X)][Y - E(Y)]\}$$

$$= D(X) + D(Y) + 2E\{[X - E(X)][Y - E(Y)]\}.$$

因为 X 与 Y 相互独立, 所以 $X - E(X)$ 与 $Y - E(Y)$ 也相互独立, 由数学期望性质得 $E\{[X - E(X)][Y - E(Y)]\} = E[X - E(X)]E[Y - E(Y)] = [E(X) - E(X)][E(Y) - E(Y)] = 0$,

从而得到 $D(X+Y) = D(X) + D(Y)$.

简单叙述为“独立随机变量和的方差等于方差之和”。

这表明“独立随机变量差的方差也等于方差之和”。

一般地: 若 $X_1, X_2, \dots, X_n$ 相互独立, 则有

$$D\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n D(X_k).$$

进一步, 若 $X_1, X_2, \dots, X_n$ 相互独立, $a_1, a_2, \dots, a_k, b$ 为常数, 则

$$D\left(\sum_{k=1}^n a_k X_k + b\right) = \sum_{k=1}^n a_k^2 D(X_k).$$

例、设 X 服从参数为 n, p 的二项分布, 即设 $X \sim B(n, p)$. 求 $D(X)$.

解: 我们知道, X 可表示为在 n 次重复独立试验中事件 A 发生的次数, p 为每次试验事件 A 发生的概率, $q=1-p$ 为每次试验事件 A 不发生的概率. 现用随机变量的分解法将随机变量 X 进行分解. 引入随机变量

$$X_i = \begin{cases} 1, & \text{在第 } i \text{ 次试验中 } A \text{ 发生,} \\ 0, & \text{在第 } i \text{ 次试验中 } A \text{ 不发生,} \end{cases} \quad i = 1, 2, \dots, n,$$

$$\text{则在 } n \text{ 次重复独立试验中事件 } A \text{ 发生的次数 } X = \sum_{i=1}^n X_i.$$

显然 $X_1, X_2, \dots, X_n$ 相互独立, 均服从(0-1)分布

$D(X_i) = pq, i=1, 2, \dots, n$, 所以方差的性质可得

$$D(X) = \sum_{i=1}^n D(X_i) = npq.$$

7. 在概率统计中, 有时需要对随机变量“标准化”, 即对任何随机变量 X , 若它的数学期望 $E(X)$, 方差 $D(X)$ 都存在, 且 $D(X) > 0$, 则称

$$Y = \frac{X - E(X)}{\sqrt{D(X)}} \text{ 为 } X \text{ 的标准化随机变量.}$$

易见 Y 是一无量纲的随机变量. 注意到 $E(X), D(X)$ 为常数, 由期望性质 2 和方差性质 2 得

$$E(Y) = E\left[\frac{X - E(X)}{\sqrt{D(X)}}\right] = \frac{E[X - E(X)]}{\sqrt{D(X)}} = \frac{E(X) - E(X)}{\sqrt{D(X)}} = 0,$$

$$D(Y) = D\left[\frac{X - E(X)}{\sqrt{D(X)}}\right] = \frac{D[X - E(X)]}{D(X)} = \frac{D(X)}{D(X)} = 1,$$

这正是标准化随机变量所具有的特征. 特别, 若 $X \sim N(\mu, \sigma^2)$, 我们已经知道 $E(X) = \mu$,

$$D(X) = \sigma^2, \text{ 则 } X \text{ 的标准化随机变量, } Y = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

第二节 随机向量

一、离散型随机向量

1、一个向量例如二维向量 (X, Y) 的横纵坐标 X, Y 都是随机变量, 这样的向量叫随机向量

2. 如果二维随机向量 (X, Y) 的全部取值数对为有限个或至多可列个, 则称随机向量 (X, Y) 为离散型的。

(1) 把数集 $\{p_{ij} | p_{ij} = P(X=x_i, Y=y_j), i, j=1, 2, \dots\}$ 做成表格, 叫 (X, Y) 的联合概率分布。

(2) 随机向量 $X = (X_1, X_2, \dots, X_n)$ 中每一个 X_i 的分布, 称为 X 关于 X_i 的**边缘分布**。

(3) 若 (X, Y) 的联合概率分布满足 $P\{X=x_i, Y=y_j\} = P(X=x_i)P(Y=y_j)$ 恒成立称 X 与 Y 独立。

例 1 某盒子中有形状相同的 2 个白球, 3 个黑球。从中一个个取球, 令

$$Y_i = \begin{cases} 1 & \text{第 } i \text{ 次取白球} \\ 0 & \text{第 } i \text{ 次取黑球} \end{cases} \quad i = 1, 2 \quad \text{分放回或不放回情形}$$

求 (Y_1, Y_2) (1) 的联合概率分布 (2) 边缘概率分布 (3) 讨论 Y_1 与 Y_2 的独立性

解: 在放回情形

联合分布

$Y_1 \backslash Y_2$	0	1
0	9/25	6/25
1	6/25	4/25

边缘分布

Y_1	0	1
P	3/5	2/5

独立

Y_2	0	1
P	3/5	2/5

$$P\{Y_1=y_i, Y_2=y_j\} = P(Y_1=y_i) P(Y_2=y_j)$$

在无放回情形

联合分布

$Y_1 \backslash Y_2$	0	1
0	3/10	3/10
1	3/10	1/10

边缘分布

Y_1	0	1
P	3/5	2/5

不独立

Y_2	0	1
P	3/5	2/5

$$P\{Y_1=y_i, Y_2=y_j\} \neq P(Y_1=y_i) P(Y_2=y_j)$$

例 2 二维随机向量 (X, Y) 的联合概率分布为:

求: (1) 常数 a 的取值; (2) $P(X \geq 0, Y \leq 1)$;

(3) $P(X \leq 1, Y \leq 1)$

解 (1) 由 $\sum p_{ij} = 1$ 得: $a = 0.1$

(2) $P(X \geq 0, Y \leq 1) = 0.1 + 0.1 + 0.2 + 0.2 = 0.6$

(3) $P(X \leq 1, Y \leq 1) = 1 - 0.25 = 0.75$

$X \backslash Y$	0	1	2
-1	0.05	0.1	0.1
0	0.1	0.2	0.1
1	a	0.2	0.05

例 3、设随机变量 X, Y 是相互独立的, 且 X, Y 等可能地取 0, 1 为值, 求随机变量 $Z = \max(X, Y)$ 的分布列。

解:

X	0	1
P	1/2	1/2

Y	0	1
P	1/2	1/2

(X, Y) 的取值数对为 $(0, 0), (0, 1), (1, 0), (1, 1)$,

$Z = \max(X, Y)$ 的取值为: 0, 1

$$P(Z=0) = P(X=0, Y=0) = P(X=0)P(Y=0) = 1/4$$

$$P(Z=1) = P(X=0, Y=1) + P(X=1, Y=0) + P(X=1, Y=1) = 3/4$$

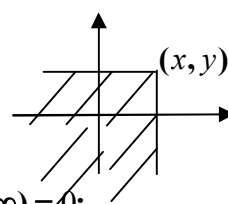
所以, Z 的分布列为

Z	0	1
P	1/4	3/4

二、随机向量的联合分布函数

1、定义函数 $F(x, y) = P\{X \leq x, Y \leq y\}$ ($x, y \in \mathbb{R}^2$) 叫**随机向量的联合分布函数**

2、二维联合分布函数区域 $\{(X, Y) | X \leq x, Y \leq y\}$ (定义域) 演示图:



3、联合分布函数性质

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

$$(3) P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

(4)如果(X,Y)为离散型随机向量,其联合概率分布为

$$P(X = x_i, Y = y_j) = p_{ij} \quad (i = 1, 2, \dots, j = 1, 2, \dots)$$

$$\text{则 } F(x, y) = P\{X \leq x, Y \leq y\} = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}$$

3. 连续型随机向量的联合概率密度

$$F(x, y) = P\{X \leq x, Y \leq y\} = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds$$

$$\text{性质 (1) } f(x, y) \geq 0, (x, y) \in \mathbb{R}^2 \quad (2) \iint_{\mathbb{R}^2} f(x, y) dx dy = 1$$

$$(3) P\{(X, Y) \in D\} = \iint_D f(x, y) dx dy$$

$$\text{特别 在 } f(x, y) \text{ 的连续点有 } \frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$$

$$\text{例 5 设 } (X, Y) \sim f(x, y) = \begin{cases} Ae^{-(2x+3y)}, & x \geq 0, y \geq 0 \\ 0, & \text{其它} \end{cases}$$

试求:(1)常数 A;(2) $P\{X < 2, Y < 1\}$; (3) $P\{(X, Y) \in D\}$,其中 D 为 $2x+3y \leq 6$.

(4) $P(X \leq x, Y \leq y)$

$$\text{解 (1) } \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} Ae^{-(2x+3y)} dx dy = \int_0^{+\infty} \int_0^{+\infty} Ae^{-2x} e^{-3y} dx dy$$

$$= A \int_0^{+\infty} e^{-2x} dx \int_0^{+\infty} e^{-3y} dy = A \left(-\frac{1}{2} e^{-2x} \right) \Big|_0^{+\infty} \left(-\frac{1}{3} e^{-3y} \right) \Big|_0^{+\infty} = A / 6 = 1, A = 6$$

$$\begin{aligned} (2) P\{X < 2, Y < 1\} &= \iint_{\{X < 2, Y < 1\}} f(x, y) dx dy = 6 \int_0^2 e^{-2x} dx \int_0^1 e^{-3y} dy = 6 \left(-\frac{1}{2} e^{-2x} \right) \Big|_0^2 \left(-\frac{1}{3} e^{-3y} \right) \Big|_0^1 \\ &= (1 - e^{-4})(1 - e^{-3}) \end{aligned}$$

$$(3) P\{(X, Y) \in D\}, \text{其中 } D \text{ 为 } 2x+3y \leq 6. \quad P\{(X, Y) \in D\} = \iint_D f(x, y) dx dy = \iint_{2x+3y \leq 6} f(x, y) dx dy$$

$$= 6 \int_0^3 e^{-2x} \left(-\frac{1}{3} e^{-3y} \right) \Big|_{\frac{2}{3}(6-x)}^0 dx = 2 \int_0^3 (e^{-2x} - e^{-6}) dx = 1 - 7e^{-6}$$

$$(4) F(x, y) = P\{X \leq x, Y \leq y\} = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds$$

所以, 当 $x \geq 0, y \geq 0$ 时

$$\begin{aligned} \int_0^x \int_0^y 6e^{-(2s+3t)} dt ds &= 6 \int_0^x e^{-2s} ds \int_0^y e^{-3t} dt = 6 \left(-\frac{1}{2} e^{-2s} \right) \Big|_0^x \left(-\frac{1}{3} e^{-3t} \right) \Big|_0^y \\ &= (1 - e^{-2x})(1 - e^{-3y}) \end{aligned}$$

$$P(X \leq x, Y \leq y) = \begin{cases} (1 - e^{-2x})(1 - e^{-3y}) & x \geq 0, y \geq 0 \\ 0 & \text{其它} \end{cases}$$

例 6 设 $(X,Y) \sim f(x,y) = \begin{cases} 4xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{其它} \end{cases}$

求 (X,Y) 的联合分布函数.

解 (1) $x < 0$, 或 $y < 0$ 时, $F(x,y) = 0$

(2) $x \geq 1, y \geq 1$ 时, $F(x,y) = 1$

(3) $0 \leq x \leq 1, 0 \leq y \leq 1$ 时,

$$F(x,y) = \int_0^x ds \int_0^y 4stdt = x^2 y^2$$

$$F(x,y) = \begin{cases} 0 & x < 0 \text{ 或 } y < 0 \\ x^2 y^2 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ x^2 & 0 \leq x \leq 1, y > 1 \\ y^2 & x > 1, 0 \leq y \leq 1 \\ 1 & x \geq 1, y \geq 1 \end{cases}$$

4、联合分布函数与边缘分布函数的关系

(1) 定义 设二维随机向量 (X, Y) 的联合分布函数为 $F(x,y)$, 则称

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

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^x f(s,t) dt ds = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f(s,t) dt \right) ds = \int_{-\infty}^x f_X(s) ds$$

$$F_Y(y) = F(+\infty, y) = P(X < +\infty, Y \leq y) = P(Y \leq y)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^y f(s,t) ds dt = \int_{-\infty}^y \left(\int_{-\infty}^{+\infty} f(s,t) ds \right) dt = \int_{-\infty}^y f_Y(t) dt$$

分别为 (X,Y) 关于 X 和 Y 的边缘分布函数.

(2) 边缘密度函数

对于连续型随机向量 $(X,Y) \sim f(x,y)$, 分量 X, Y 的密度函数称为边缘密度函数。

已知联合密度函数, 容易求出边缘密度函数。

$$f_X(x) = f_1(x) = \int_{-\infty}^{+\infty} f(x,y) dy, f_Y(y) = f_2(y) = \int_{-\infty}^{+\infty} f(x,y) dx$$

事实上, (1) $f_1(x) \geq 0$, (2) 若 $a < b$, 则

$$P\{a < X < b\} = P\{a < X < b, -\infty < Y < +\infty\} = \int_a^b dx \int_{-\infty}^{+\infty} f(x,y) dy = \int_a^b f_1(x) dx$$

所以, $f_1(x)$ 是 X 的概率密度, 同理可证 $f_2(y)$.

5、两个随机变量的关系

随机变量的相互独立性

定理 1 随机变量 X 与 Y 是相互独立的 $\Leftrightarrow \begin{cases} p_{ij} = p_i \cdot p_j & \text{离散型} \\ f(x,y) = f_1(x) f_2(y) & \text{连续型} \end{cases}$

定理 2 若 X 与 Y 相互独立, 则它们的连续函数 $g(X)$ 与 $h(Y)$ 也相互独立。

特别有 $aX+b$ 与 $cY+d$ 相互独立。

例 1、已知随机向量 (X,Y) 的联合密度为 $f(x,y) = \begin{cases} e^{-x-y}, & x > 0, y > 0; \\ 0, & \text{其他。} \end{cases}$

(1)问 X 与 Y 是否独立? (2)求概率 $P\{X<Y\}$.

$$\text{解 (1)} \quad f_1(x) = \begin{cases} \int_0^{+\infty} e^{-(x+y)} dy = e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}, f_2(y) = \begin{cases} \int_0^{+\infty} e^{-(x+y)} dx = e^{-y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

$f(x, y) = f_1(x)f_2(y)$, 所以, X, Y 独立.

$$(2) P(X < Y) = \iint_{x < y} f(x, y) dx dy = \int_0^{+\infty} dx \int_x^{+\infty} e^{-(x+y)} dy = \frac{1}{2}$$

例 2、设二维随机变量 (X, Y) 的概率密度为

$$f(x, y) = \begin{cases} e^{-y}, & 0 < x < y \\ 0, & \text{其他} \end{cases} \quad (1) \text{ 求随机变量 X 的密度函数; } (2) \text{ 求概率 } P\{X+Y \leq 1\}.$$

解 (1) $x \leq 0$ 时, $f_1(x) = 0$;

$$x > 0 \text{ 时, } f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_x^{+\infty} e^{-y} dy = e^{-x}$$

$$\text{所以, } f_1(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$(2) P\{X+Y \leq 1\} = \int_0^{1/2} dx \int_x^{1-x} e^{-y} dy = 1 + e^{-1} - 2e^{-\frac{1}{2}}$$

例 3、已知 X, Y 的分布函数为 $F(x, y) = A(B + \arctan \frac{x}{2})(C + \arctan \frac{y}{3})$.

求: (1) A, B, C (2) (X, Y) 的概率密度 (3) 边缘概率密度和边缘分布函数
(4) 讨论的独立性

$$(1) A = \frac{1}{\pi^2}, B = \frac{\pi}{2}, C = \frac{\pi}{2};$$

$$(2) (X, Y) \text{ 的概率密度 } f(x, y) = \frac{6}{\pi^2(4+x^2)(9+y^2)};$$

$$(3) f_X(x) = \frac{2}{\pi(4+x^2)}, F_X(x) = \frac{1}{\pi}(\frac{\pi}{2} + \arctan \frac{x}{2});$$

$$f_Y(y) = \frac{3}{\pi(9+y^2)}, F_Y(y) = \frac{1}{\pi}(\frac{\pi}{2} + \arctan \frac{y}{3});$$

$$(4) f(x, y) = f_X(x) f_Y(y) \text{ or } F(x, y) = F_X(x) F_Y(y) \text{ 独立.}$$

三、常见的二维连续型随机向量

1、均匀分布 若二维随机向量 (X, Y) 的联合概率密度为 $f(x, y) = \begin{cases} \frac{1}{S_D} & (x, y) \in D \\ 0 & \text{其它} \end{cases}$

其中: D 为可度量的平面区域, S_D 为区域 D 的面积. 则称 (X, Y) 服从区域 D 上的均匀分布.

对于 D 中任意可度量区域 G 有 $P\{(x,y) \in G\} = \iint_G f(x,y)dxdy = \iint_G \frac{1}{S_D} dxdy = \frac{S_G}{S_D}$

其中:SG 为区域 G 的面积.

例 1、 设随机向量(X,Y)服从区域 D 上的均匀分布,

其中 $D=\{(x,y) \mid x^2+y^2 \leq 1\}$,求 X,Y 的边缘密度函数 $f_1(x)$ 和 $f_2(y)$.

解 (1)由题意得: $f(x,y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1 \\ 0 & \text{其它} \end{cases}$

$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy$ 当 $|x|>1$ 时, $f(x,y)=0$,所以, $f_1(x)=0$

当 $|x| \leq 1$ 时, $f_1(x) = [\int_{-\infty}^{-\sqrt{1-x^2}} + \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} + \int_{\sqrt{1-x^2}}^{+\infty}] f(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2}$

所以, $f_1(x) = \begin{cases} \frac{2}{\pi} \sqrt{1-x^2} & |x| \leq 1 \\ 0 & |x| > 1 \end{cases}$ 同理, $f_2(y) = \begin{cases} \frac{2}{\pi} \sqrt{1-y^2} & |y| \leq 1 \\ 0 & |y| > 1 \end{cases}$ 不独立

注均匀分布的边缘密度不再是一维均匀分布

2、二维正态分布

定义 如果(X,Y)的联合密度函数为

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

其中 $-\infty < \mu_1, \mu_2 < +\infty, \sigma_1^2 > 0, \sigma_2^2 > 0, |\rho| < 1$,

则称(X,Y)服从参数为 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 的二维正态分布,简记为

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

可以证明 若 $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 则 X,Y 的边缘概率密度分别为

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

即 二维正态分布(X,Y)的边缘概率密度是一维正态分布.

由此可知随机向量的联合概率密度完全决定了它的边缘概率密度,反之不一定成立.

例 2、 设二维随机向量(X,Y)的联合概率密度为

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

求(X,Y)关于 X,Y 的边缘概率密度.

解 $f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} (1+xy)dy$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{同理可得 } f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

(X,Y)的边缘概率密度为一维正态分布.

但边缘概率密度为一维正态分布的二维随机向量不一定是二维正态分布.

例 3、 设二维随机向量 $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, $f(x,y)$ 为其联合密度函数, 证明 X 与 Y 独立的充要条件是 $\rho=0$

证明 由题意得 $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$, $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$

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

充分性 将 $\rho=0$ 代入 $f(x,y)$ 即得 $f(x,y)=f_X(x)f_Y(y)$.

必要性 若 X 和 Y 相互独立, 则 $f(x,y)=f_X(x)f_Y(y)$,

特别取 $x = \mu_1, y = \mu_2$, 得 $\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} = \frac{1}{2\pi\sigma_1\sigma_2}$ 即 $\rho=0$

三、随机向量的数字特征

1、随机向量的数学期望和方差

定义 设 $(X_1, X_2, \dots, X_n)$ 是 n 维随机向量, 且每个分量的数学期望和方差都存在;

称数值向量 $(EX_1, EX_2, \dots, EX_n)$ 为 $(X_1, X_2, \dots, X_n)$ 的数学期望, 简称期望;

称数值向量 $(DX_1, DX_2, \dots, DX_n)$ 为 $(X_1, X_2, \dots, X_n)$ 的方差.

2、随机变量函数的数学期望

设 $g(X,Y)$ 为随机变量 X,Y 的函数, $E[g(X,Y)]$ 存在

(1) 若 (X,Y) 为离散型随机向量, $P(X=x_i, Y=y_j)=p_{ij}$, $(i,j=1,2,\dots)$, 则

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

(2) 若 (X,Y) 为连续型随机向量, $(X,Y) \sim f(x,y)$, 则

$$E[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f(x,y) dx dy$$

3、数学期望, 方差的性质

(1) **离散型** 设 (X,Y) 联合分布为 $P(X=x_i, Y=y_j)=p_{ij}$, $(i,j=1,2,\dots)$

$$EX = \sum_i \sum_j x_i p_{ij} = \sum_i x_i p_{i.}, EY = \sum_i \sum_j y_j p_{ij} = \sum_j y_j p_{.j};$$

$$E(X \pm Y) = \sum_i \sum_j (x_i \pm y_j) p_{ij} = \sum_i \sum_j x_i p_{ij} \pm \sum_i \sum_j y_j p_{ij} = EX \pm EY$$

$$DX = E(X - EX)^2 = \sum_i (x_i - EX)^2 p_{i.} = \sum_i \sum_j (x_i - EX)^2 p_{ij}$$

若 (X,Y) 是二维随机向量, 且 X 与 Y 独立, 则 $E(XY)=EXEY$

(2) **连续型** 设 $(X,Y) \sim f(x,y)$, 则

$$EX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x,y) dx dy = \int_{-\infty}^{+\infty} xf_X(x) dx, EY = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yf(x,y) dx dy = \int_{-\infty}^{+\infty} yf_Y(y) dy$$

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (xy)f(x,y)dxdy$$

若 X,Y 相互独立则

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (xy)f_X(x)f_Y(y)dxdy = \int_{-\infty}^{+\infty} xf_X(x)dx \cdot \int_{-\infty}^{+\infty} yf_Y(y)dy = E(X)E(Y)$$

$$DX = E(X - EX)^2 = \int_{-\infty}^{+\infty} (x - EX)^2 f_X(x)dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - EX)^2 f(x,y)dxdy$$

注 性质可推广到有限个随机变量情形.

例 1 设(X,Y)的联合概率分布为

$\begin{smallmatrix} X \backslash Y \\ \end{smallmatrix}$	0	1	2	3
1	0	3/8	3/8	0
3	1/8	0	0	1/8

求 EX,EY,E(XY).

解

X	1	3
P	3/4	1/4

Y	0	1	2	3
P	1/8	1/8	3/8	1/8

XY	0	1	2	3	6	9
P	1/8	3/8	3/8	0	0	1/8

所以 EX=3/2, EY=3/2

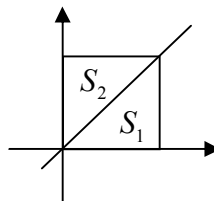
$$\begin{aligned} E(XY) &= (1 \times 0) \times 0 + (1 \times 1) \times \frac{3}{8} + (1 \times 2) \times \frac{3}{8} + (1 \times 3) \times 0 \\ &\quad + (3 \times 0) \times \frac{1}{8} + (3 \times 1) \times 0 + (3 \times 2) \times 0 + (3 \times 3) \times \frac{1}{8} = \frac{9}{4} \end{aligned}$$

例 2. 设(X,Y)服从 $[0,a] \times [0,a]$ 上的均匀分布,求 $E|X-Y|$.

解 由题意得 $(X,Y) \sim f(x,y) = \begin{cases} \frac{1}{a^2} & 0 \leq x \leq a, 0 \leq y \leq a \\ 0 & \text{其它} \end{cases}$

根据 $E[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$ 得

$$\begin{aligned} E|X-Y| &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x-y| f(x,y)dxdy = \int_0^a \int_0^a |x-y| \frac{1}{a^2} dxdy \\ &= \iint_{S_1} |x-y| \frac{1}{a^2} dxdy + \iint_{S_2} |x-y| \frac{1}{a^2} dxdy \\ &= 2 \int_0^a dx \int_0^x (x-y) \frac{1}{a^2} dy = \frac{a}{3} \end{aligned}$$



4. 协方差

定义对两个随机向量(X,Y),若 $E(X-EX)(Y-EY)$ 存在,则称 $\text{cov}(X,Y)=E(X-EX)(Y-EY)$ 为 X 和 Y

的协方差。

特别, 若 $X=Y$, 则 $\text{cov}(X, X) = E(X - EX)^2 = DX$

因此, 方差是协方差的特例, 协方差刻画两个随机变量之间的“某种”关系。

可以证明 若 (X, Y) 服从二维正态分布, 即

若 $(X, Y) \sim N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$, 则 $\text{cov}(X, Y) = \rho\sigma_1\sigma_2$

性质 ① $\text{cov}(X, Y) = EXY - EXEY$

推论 $D(X \pm Y) = DX + DY \pm 2\text{cov}(X, Y)$

② $\text{cov}(X, Y) = \text{cov}(Y, X)$

③ $\text{cov}(aX, bY) = ab\text{cov}(X, Y)$

④ $\text{cov}(X_1 + X_2, Y) = \text{cov}(X_1, Y) + \text{cov}(X_2, Y)$

证明 ④ $\text{cov}(X_1 + X_2, Y) = E[(X_1 + X_2)Y] - E(X_1 + X_2)EY$

$$= E(X_1Y + X_2Y) - (EX_1 + EX_2)EY$$

$$= E(X_1Y) + E(X_2Y) - (EX_1)(EY) - (EX_2)(EY)$$

$$= E(X_1Y) - (EX_1)(EY) + E(X_2Y) - (EX_2)(EY)$$

$$= \text{cov}(X_1, Y) + \text{cov}(X_2, Y)$$

5. 相关系数

定义 设随机变量 X 和 Y 的方差为正值, 称 $\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}}$ 为 X 与 Y 的相关系数

注 $\rho_{XY} = (EXY - EXEY) / \sqrt{DX}\sqrt{DY}$

$$D(X \pm Y) = DX + DY \pm 2\text{cov}(X, Y) = DX + DY \pm 2\rho_{XY}\sqrt{DX}\sqrt{DY}$$

若 (X, Y) 服从二维正态分布, 则 $\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho$

例 1、设 (X, Y) 服从二维正态分布, 且 $X \sim N(1, 9)$, $Y \sim N(0, 16)$,

若 $\rho_{XY} = -\frac{1}{2}$, $Z = \frac{X}{3} + \frac{Y}{2}$, 求 ρ_{XZ} .

$$\begin{aligned} \text{解: } \text{cov}(X, Z) &= \text{cov}\left(X, \frac{X}{3} + \frac{Y}{2}\right) = \text{cov}\left(X, \frac{X}{3}\right) + \text{cov}\left(X, \frac{Y}{2}\right) \\ &= \frac{1}{3}\text{cov}(X, X) + \frac{1}{2}\text{cov}(X, Y) = \frac{1}{3}DX + \frac{1}{2}\rho_{XY}\sqrt{DX}\sqrt{DY} \\ &= \frac{1}{3} \times 9 + \frac{1}{2} \times \left(-\frac{1}{2}\right) \times 3 \times 4 = 0 \end{aligned}$$

$$\text{所以 } \rho_{XZ} = \frac{\text{cov}(X, Z)}{\sqrt{DX}\sqrt{DZ}} = 0$$

性质 (1) $|\rho_{XY}| \leq 1$ (2) 若 $Y = aX + b$, 则 $\begin{cases} a > 0 \text{ 时, } \rho_{XY} = 1 \\ a < 0 \text{ 时, } \rho_{XY} = -1 \end{cases}$

(3) $\rho_{XY} = \pm 1 \Leftrightarrow P(Y = aX + b) = 1$ (4) 若 X, Y 相互独立, 则 $\rho_{XY} = 0$

$$\begin{aligned}
\text{证明(2)} \quad \text{cov}(X,Y) &= E(X-EX)(Y-EY) \\
&= E(X-EX)[(aX+b)-E(aX+b)] \\
&= E(X-EX)(aX-aEX) = aE(X-EX)^2 = aDX \\
DY &= D(aX+b) = a^2DX
\end{aligned}$$

$$\text{故 } \rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{DX}\sqrt{DY}} = \frac{aDX}{\sqrt{DX}\sqrt{a^2DX}} = \frac{a}{|a|} = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

注意 $|\rho_{XY}|$ 的大小反映了 X,Y 之间线性关系的密切程度:

$\rho_{XY}=0$ 时, X,Y 之间无线性关系;

$|\rho_{XY}|=1$ 时, X,Y 之间具有线性关系.

由此定义:

$\rho_{XY} \neq 0$, X,Y 相关 ($\rho_{XY}=1$, X,Y 完全正线性相关),

$\rho_{XY}=0$, X,Y 不相关 ($\rho_{XY}=-1$, X,Y 完全负线性相关)

显然

$$X,Y \text{ 独立} \Rightarrow X \text{ 与 } Y \text{ 不相关} \Leftrightarrow \rho_{XY} = 0 \Leftrightarrow \text{cov}(X,Y) = 0$$

$$\Leftrightarrow EXY = EXEY \Leftrightarrow D(X \pm Y) = DX + DY$$

注 X,Y 不相关, 不一定有 X,Y 独立.

只有(X,Y)服从二维正态分布, 则 X,Y 独立 $\nleftrightarrow$ X,Y 不相关.

$$\text{例 1、 随机向量}(X,Y) \sim f(x,y) = \begin{cases} \frac{1}{8}(x+y) & 0 < x < 2 \quad 0 < y < 2 \\ 0 & \text{其它} \end{cases}$$

求 X,Y 的相关系数 ρ_{XY} .

$$\text{解 据定理 2, } E[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy$$

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

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

$$\text{由对称性得 } E(Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yf(x,y)dxdy = 5/3$$

$$\text{由对称性得 } E(Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} yf(x,y)dxdy = 7/6, EY^2 = 5/3$$

$$\text{所以, } DX = 11/36, \quad DY = 11/36$$

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

$$\rho_{XY} = (EXY - EXEY) / \sqrt{DX}\sqrt{DY} = -\frac{1}{11} \neq 0 \text{ 不独立}$$

例 2、设随机向量(X,Y)服从区域 D 上的均匀分布,
其中 $D=\{(x,y), x^2+y^2 \leq 1\}$, 求 X,Y 的边缘密度函数 $f_1(x)$ 和 $f_2(y)$,
问 X,Y 相关吗?独立吗?

解 (1)由题意得: $f(x,y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1 \\ 0 & \text{其它} \end{cases}, f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy$

$$f_1(x) = \begin{cases} \frac{2}{\pi} \sqrt{1-x^2} & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \text{同理, } f_2(y) = \begin{cases} \frac{2}{\pi} \sqrt{1-y^2} & |y| \leq 1 \\ 0 & |y| > 1 \end{cases} \text{不独立}$$

$$0 = E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (xy)f(x,y)dxdy = EXEY = 0 \times 0 = 0 \text{ 不相关}$$

例 3、(X,Y)的联合分布为:

求相关系数 ρ_{XY} , 并判断 X,Y 是否相关, 是否独立.

解: $E(X) = \sum_i \sum_j x_i p_{ij} = 0, E(X) = \sum_i x_i p_{i.} = 0$

$$E(X^2) = \sum_i \sum_j x_i^2 p_{ij} = 3/4$$

$$\{\text{或者 } E(X^2) = \sum_i x_i^2 p_{i.} = 3/4\}$$

由对称性得 $EY=EX=0 \quad EY^2=EX^2=3/4$

$$\text{另外 } E(XY) = \sum_i \sum_j (x_i y_j) p_{ij} = 1/8 - 1/8 - 1/8 + 1/8 = 0$$

所以 $\text{Cov}(X,Y) = EXY - EXEY = 0$ 即 X 与 Y 不相关

亦即 $\rho_{XY} = 0$ 另一方面

$$P(X=-1, Y=-1) = 1/8 \neq 0$$

$$P(X=-1)P(Y=-1) = (3/8) \times (3/8)$$

所以 X 与 Y 不独立

四、随机变量函数的分布

1、离散型随机变量函数的概率分布

例 1 设随机变量 X 的概率分布如下,
 $Y=2X+1, Z=X^2$, 求 Y,Z 的概率分布.

解 (1)Y 的对应取值为 -1, 1, 3, 5,

$$P(Y=-1) = P(2X+1=-1) = P(X=-1) = 0.2$$

$$P(Y=1) = P(X=0) = 0.3, \quad P(Y=3) = P(X=1) = 0.1, \quad P(Y=5) = P(X=2) = 0.4,$$

所以 Y 的概率分布为

(2)Z 的取值为 0, 1, 4,

$$P(Z=0) = P(X=0) = 0.3,$$

$$P(Z=1) = P(X=1) + P(X=-1) = 0.3,$$

$$P(Z=4) = P(X=2) = 0.4$$

所以 Z 的概率分布为:

X	-1	0	1	2
P	0.2	0.3	0.1	0.4

Y	-1	1	3	5
P	0.2	0.3	0.1	0.4

Z	0	1	4
P	0.3	0.3	0.4

2、连续型随机变量函数的概率密度函数

例 2 设随机变量 $X \sim f_X(x) = \frac{1}{\pi(1+x^2)}, Y=2X+1$, 求随机变量 Y 的概率密度函数 $f_Y(y)$.

解 (1)求 Y 的分布函数 $F_Y(y)$: $F_Y(y) = P(Y \leq y) = P(2X+1 \leq y) = P(X \leq \frac{y-1}{2}) = F_X(\frac{y-1}{2})$

(2) 在 $F_Y(y) = F_X(\frac{y-1}{2})$ 两边对 y 求导得

$$f_Y(y) = f_X(\frac{y-1}{2}) (\frac{y-1}{2})' = \frac{1}{2} f_X(\frac{y-1}{2}) = \frac{1}{2} \frac{1}{\pi [1 + (\frac{y-1}{2})^2]} = \frac{2}{\pi [4 + (y-1)^2]}$$

定理 1 设 $X \sim f_X(x)$, $y=g(x)$ 是 x 的单调可导函数, 其导数不为 0, 值域为 (a, b) , $-\infty < a < b < +\infty$, 记 $x=h(y)$ 为 $y=g(x)$ 的反函数, 则 $Y=g(X)$ 的概率密度函数为:

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot |h'(y)| & a < y < b \\ 0 & \text{其它} \end{cases}$$

设 $y=g(x) \uparrow \quad F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$

$$f_Y(y) = F_Y'(y) = f_X(h(y))(h(y))'$$

特别: 对随机变量 X 的线性函数有以下定理,

定理 2 设随机变量 $X \sim F_X(x)$, $Y=kX+b(k \neq 0)$, 则 Y 的概率密度为

$$f_Y(y) = \frac{1}{|k|} f_X(\frac{y-b}{k})$$

例如 设 X 为连续型随机变量, $X \sim F_X(x)$, $Y = -4X+3$,

$$\text{则 } Y \text{ 的密度函数为 } f_Y(y) = \frac{1}{4} f_X(\frac{y-3}{-4})$$

例 3 设 $X \sim N(\mu, \sigma^2)$, $Y = \frac{X-\mu}{\sigma}$, 则 $Y \sim N(0, 1)$.

证明: $Y = \frac{1}{\sigma} X - \frac{\mu}{\sigma} \Rightarrow x = \sigma y + \mu \uparrow$

$$f_Y(y) = (\sigma y + \mu)' f_X(\sigma y + \mu) = \sigma f_X(\sigma y + \mu)$$

$$= \sigma \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\sigma y + \mu - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \phi(y)$$

所以 $Y \sim N(0, 1)$

例 4、设 $X \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$, 求 $Y = \tan X$ 的概率密度。

$$\text{解: } y = \tan x \uparrow \Rightarrow x = \arctan y, x' = \frac{1}{1+y^2}, f_Y(y) = \frac{1}{\pi(1+y^2)}$$

例 5、设 $X \sim E(1)$, 求 $Y = \ln X$ 的概率密度。

$$\text{解: } y = \ln x \uparrow \Rightarrow x = e^y, x' = e^y, f_Y(y) = e^{-e^y} \cdot e^y$$

例 6、设 $X \sim U(0, \pi)$, 求 $Y = \sin X$ 的概率密度。

$$\text{解 } y = \sin x \Rightarrow x = \begin{cases} \arcsin y & 0 \leq y \leq 1 \\ \pi - \arcsin y & 0 \leq y \leq 1 \end{cases}, x' = \frac{\pm 1}{\sqrt{1-y^2}},$$

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

例 7 设 X 与 Y 同分布, X 的概率密度为 $f(x) = \begin{cases} \frac{3}{8}x^2, & 0 < x < 2 \\ 0, & \text{其他} \end{cases}$

(1) 已知事件 $A=\{X>\alpha\}$ 和 $B=\{Y>\alpha\}$ 独立,且 $P(A+B)=3/4$.

求常数 α ; (2) 求 $E(1/X^2)$.

解 (1)由已知得: $P(A)=P(B)$,A,B 独立,所以,

$$P(A+B)=P(A)+P(B)-P(A)P(B)=2P(A)-[P(A)]^2=3/4$$

故 $P(A)=1/2, 0<\alpha<2$,

$$P(X>\alpha)=\int_{\alpha}^2 \frac{3}{8}x^2dx=1-\frac{\alpha^3}{8}=1/2 \text{ 所以, } \alpha=\sqrt[3]{4}$$

$$(2)E(1/X^2)=\int_0^2 \frac{1}{x^2} \frac{3}{8}x^2dx=3/4$$

$$\text{例 8. 设 } Y \sim f(y)=\begin{cases} \frac{y}{a^2}e^{-\frac{y^2}{2a^2}}, & y>0 \\ 0, & y<0 \end{cases}$$

求 $Z=1/Y$ 的数学期望 $E(Z)$.

$$\text{解 } EZ=E(1/Y)=\int_{-\infty}^{+\infty} \frac{1}{y} f(y)dy=\int_0^{+\infty} \frac{1}{a^2} e^{-\frac{y^2}{2a^2}} dy=\frac{1}{a^2} \int_0^{+\infty} e^{-\left(\frac{y}{\sqrt{2}|a|}\right)^2} dy$$

$$=\frac{1}{a^2} \int_0^{+\infty} \sqrt{2}|a| e^{-\left(\frac{y}{\sqrt{2}|a|}\right)^2} d\left(\frac{y}{\sqrt{2}|a|}\right)$$

$$\text{令 } u=\frac{y}{\sqrt{2}|a|}, \quad EZ=\frac{\sqrt{2}|a|}{a^2} \int_0^{+\infty} e^{-u^2} du=\frac{\sqrt{2}|a|}{a^2} \frac{\sqrt{\pi}}{2}=\frac{\sqrt{2\pi}}{2|a|}$$

3、离散型随机向量函数的分布

例 1、 设(X,Y)的联合概率分布为:

求:(1) $X+Y, XY, X^2+Y^2$ 的概率分布.

解:

$X \backslash Y$	0	1	2
-1	0.4	0.1	0.1
0	0.1	0.2	0.1

(X, Y)	(-1,0)	(-1,1)	(-1,2)	(0,0)	(0,1)	(0,2)
$X+Y$	-1	0	1	0	1	2
XY	0	-1	-2	0	0	0
X^2+Y^2	1	2	5	0	1	4
P	0.4	0.1	0.1	0.1	0.2	0.1

$X+Y$	-1	0	1	2
P	0.4	0.2	0.3	0.1

XY	-2	-1	0
P	0.1	0.1	0.8

X^2+Y^2	0	1	2	5	4
P	0.1	0.6	0.1	0.1	0.1

例 2 设随机变量 X_1 与 X_2 相互独立, 分别服从二项 $B(n_1, p)$

分布和 $B(n_2, p)$, 求 $Y=X_1+X_2$ 的概率分布.

解 依题知 $X+Y$ 的可能取值为 $0, 1, 2, \dots, n_1+n_2$, 因此对于 k ($k=0, 1, 2, \dots, n_1+n_2$), 由独立性有

$$\begin{aligned} P(Y=k) &= \sum_{k_1+k_2=k} P(X_1=k_1, X_2=k_2) = \sum_{k_1+k_2=k} C_{n_1}^{k_1} p^{k_1} (1-p)^{n_1-k_1} C_{n_2}^{k_2} p^{k_2} (1-p)^{n_2-k_2} \\ &= \sum_{k_1+k_2=k} C_{n_1}^{k_1} C_{n_2}^{k_2} p^k (1-p)^{n_1+n_2-k} \end{aligned}$$

$$\text{由 } \sum_{k_1+k_2=k} C_{n_1}^{k_1} C_{n_2}^{k_2} = C_{n_1+n_2}^k \text{ 得 } P(Y=k) = C_{n_1+n_2}^k p^k (1-p)^{n_1+n_2-k}$$

于是 $Y=X_1+X_2$ 服从二项分布 $B(n_1+n_2, p)$

若 X 与 Y 相互独立, $X \sim B(n_1, p)$, $Y \sim B(n_2, p)$,

则 $X+Y \sim B(n_1+n_2, p)$ 二项分布的可加性

类似可得:

若 X, Y 相互独立, $X \sim P(\lambda_1), Y \sim P(\lambda_2)$,

则 $X+Y \sim P(\lambda_1+\lambda_2)$ Poisson 分布的可加性

4、随机变量函数的概率密度函数求法----分布函数法

设随机向量 (X, Y) 的联合密度函数为 $f(x, y)$, 记 $Z=g(X, Y)$.

$$(1) \text{ 求 } Z \text{ 的分布函数 } F(z) = P(Z \leq z) = P(g(X, Y) \leq z) = \iint_{g(x, y) \leq z} f(x, y) dx dy$$

(2) 对 $F(z)$ 求导即得 Z 的概率密度函数 $f(z)$.

例 1、 设随机变量 X 与 Y 独立, 概率密度函数为

$$f_X(x) = \begin{cases} \frac{2}{\sqrt{\pi}} e^{-x^2} & x > 0 \\ 0 & \text{其他} \end{cases}, f_Y(y) = \begin{cases} \frac{2}{\sqrt{\pi}} e^{-y^2} & y > 0 \\ 0 & \text{其他} \end{cases}$$

求 $Z = \sqrt{X^2 + Y^2}$ 的概率密度函数.

$$\text{解 } (X, Y) \text{ 的联合密度函数为 } f(x, y) = \begin{cases} \frac{4}{\pi} e^{-(x^2+y^2)} & x > 0, y > 0 \\ 0 & \text{其他} \end{cases}$$

$$(1) z \leq 0 \text{ 时, } F_Z(z) = P(Z \leq z) = P(\sqrt{X^2 + Y^2} \leq z) = 0, \quad f_Z(z) = 0$$

(2) $z > 0$ 时

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(\sqrt{X^2 + Y^2} \leq z) \\ &= \iint_{\sqrt{x^2+y^2} \leq z} \frac{4}{\pi} e^{-(x^2+y^2)} dx dy = \frac{4}{\pi} \int_0^{\frac{\pi}{2}} d\theta \int_0^z e^{-r^2} r dr = -e^{-z^2} + 1 \\ f_Z(z) &= F'_Z(z) = 2ze^{-z^2} \end{aligned}$$

$$\text{所以, } f_Z(z) = \begin{cases} 2ze^{-z^2} & z > 0 \\ 0 & z \leq 0 \end{cases}$$

例 2 设随机向量(X,Y)的概率密度为.

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & x > 0, y > 0 \\ 0 & \text{其他} \end{cases}, \text{求 } Z = X + Y \text{ 的概率密度}$$

解:(1)当 $z \leq 0$ 时, $F_Z(z) = 0, f_Z(z) = 0$

(2)当 $z > 0$ 时, $F_Z(z) = P(Z \leq z) = P(X + Y \leq z)$

$$= \iint_{x+y \leq z} 2e^{-(x+2y)} dx dy = \int_0^z e^{-x} dx \int_0^{z-x} 2e^{-2y} dy = 1 - 2e^{-z} + e^{-2z}$$

$$f_Z(z) = F'_Z(z) = 2ze^{-z} - 2e^{-2z}$$

$$\text{综上, } f_Z(z) = \begin{cases} 0 & z \leq 0 \\ 2e^{-z}(1 - e^{-z}) & z > 0 \end{cases}$$

* 设随机变量 X 和 Y 相互独立, 概率密度函数分别为 $f_1(x)$ 和 $f_2(y)$, 求 $Z = X + Y$ 的概率密度函数.

$$\text{解 } F_Z(z) = P\{X + Y \leq z\} = \iint_{x+y \leq z} f(x,y) dx dy = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f(x,y) dy \right) dx$$

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx = \int_{-\infty}^{+\infty} \overset{\text{独立}}{f_1(x)} f_2(z-x) dx$$

例 3 设随机向量(X,Y)的概率密度为.

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & x > 0, y > 0 \\ 0 & \text{其他} \end{cases}, \text{求 } Z = X + Y \text{ 的概率密度}$$

$$\text{解: } f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx = \begin{cases} 0 & z < 0 \\ \int_0^z e^{-[x+2(z-x)]} dx & z \geq 0 \end{cases} = \begin{cases} 0 & z < 0 \\ 2e^{-z} - 2e^{-2z} & z \geq 0 \end{cases}$$

例 4、 设随机变量 X 和 Y 相互独立, 且均服从标准正态分布 $N(0,1)$, 求 $Z = X + Y$ 的概率密度函数.

解 由题意得 $f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, f_2(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, 因 X 和 Y 相互独立, 故

$$f_Z(z) = \int_{-\infty}^{+\infty} f_1(x) f_2(z-x) dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} e^{-\frac{(z-x)^2}{2}} dx = \frac{1}{2\pi} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} e^{-(x-\frac{z}{2})^2} dx$$

$$\text{令 } \frac{t}{\sqrt{2}} = x - \frac{z}{2}$$

$$f_Z(z) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}}$$

两个独立正态分布随机变量的和仍服从正态分布.

即:若 $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, X, Y 独立, 则 $X+Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

正态分布的可加性

推论 有限个独立的正态分布的线性函数仍服从正态分布.

即:若 $X_i \sim N(\mu_i, \sigma_i^2)$, ($i=1, 2, \dots, n$), $X_1, X_2, \dots, X_n$ 相互独立,

实数 $a_1, a_2, \dots, a_n$ 不全为零, 则 $\sum_{i=1}^n a_i X_i \sim N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$

特别, 若 $X_1, X_2, \dots, X_n$ 独立同服从正态分布 $N(\mu, \sigma^2)$, 记: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, 则 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

例 5 设随机变量 X 与 Y 独立, 概率密度函数为

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

求 $Z = 2X + Y$ 的概率密度 $f_Z(z)$.

$$\text{解: } f_{2X}(x) = \begin{cases} \frac{1}{2} & 0 < x < 2 \\ 0 & \text{其他} \end{cases}, \quad f_Y(y) = \begin{cases} e^{-y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{2X}(x) f_Y(z-x) dx = \begin{cases} 0 & z \leq 0 \\ \int_0^z \frac{1}{2} e^{-(z-x)} dx = \frac{1}{2} (1 - e^{-z}) & 0 < z \leq 2 \\ \int_0^2 \frac{1}{2} e^{-(z-x)} dx = \frac{1}{2} (e^2 - 1) e^{-z} & z > 2 \end{cases}$$

* 设随机变量 X 和 Y 相互独立, 概率密度函数分别为 $f_1(x)$ 和 $f_2(y)$, 求 $Z = X - Y$ 的概率密度函数.

$$F_Z(z) = P\{X - Y \leq z\} = \iint_{x-y \leq z} f(x, y) dx dy = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{y+z} f(x, y) dx \right) dy$$

$$f_Z(z) = \int_{-\infty}^{+\infty} f(y+z, y) dy = \int_{-\infty}^{+\infty} f_1(y+z) f_2(y) dy$$

例 6、设 (X, Y) 的概率密度为 $f(x, y) = \begin{cases} e^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{其他} \end{cases}$, 求 $Z = X - Y$ 的概率密度

$$f_Z(z) = \int_{-\infty}^{+\infty} f(y+z, y) dy = \begin{cases} \int_{-z}^{+\infty} e^{-(y+z+y)} dy = \frac{e^{-z}}{2} & z < 0 \\ \int_0^{+\infty} e^{-(y+z+y)} dy = \frac{e^{-z}}{2} & z \geq 0 \end{cases} = \frac{e^{-|z|}}{2}$$

* 设随机变量 X 和 Y 相互独立, 概率密度函数分别为 $f_1(x)$ 和 $f_2(y)$, 求 $Z = X / Y$ 的概率密度函数.

数.

$$F_Z(z) = P\{X/Y \leq z\} = \iint_{x/y \leq z} f(x, y) dx dy = \int_{-\infty}^0 \left(\int_{yz}^{+\infty} f(x, y) dx \right) dy + \int_0^{+\infty} \left(\int_{-\infty}^{yz} f(x, y) dx \right) dy$$

$$f_Z(z) = \int_{-\infty}^{+\infty} |y| f(yz, y) dy = \int_{-\infty}^{+\infty} |y| f_1(yz) f_2(y) dy$$

例 7、设随机变量 X 与 Y 独立，同服从 N(0,1), 求 Z=X/Y 的概率密度.

$$\text{解: } f_Z(z) = \int_{-\infty}^{+\infty} |y| f(yz, y) dy = \int_{-\infty}^{+\infty} |y| f_1(yz) f_2(y) dy$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |y| e^{-\frac{y^2}{2}} e^{-\frac{(yz)^2}{2}} dy = \frac{1}{\pi} \int_0^{+\infty} y e^{-\frac{(1+z^2)y^2}{2}} dy = \frac{1}{\pi(1+z^2)}$$

$$(-\infty < z < +\infty)$$

* 设随机变量 X 和 Y 相互独立, 概率密度函数分别为 $f_1(x)$ 和 $f_2(y)$, 求 $Z=XY$ 的概率密度函数.

$$\text{解 } F_Z(z) = P\{XY \leq z\} = \iint_{xy \leq z} f(x, y) dx dy = \int_{-\infty}^0 \left(\int_{z/x}^{+\infty} f(x, y) dx \right) dy + \int_0^{+\infty} \left(\int_{-\infty}^{z/x} f(x, y) dx \right) dy$$

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx = \int_{-\infty}^{+\infty} \frac{1}{|x|} f_1(x) f_2(\frac{z}{x}) dx, \text{ 同理}$$

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|y|} f(\frac{z}{y}, y) dy = \int_{-\infty}^{+\infty} \frac{1}{|y|} f_1(\frac{z}{y}) f_2(y) dy$$

例 8 设随机变量 X 与 Y 独立，同服从 U(0,1), 求 Z=XY 的概率密度.

$$\text{解: } f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx = \int_{-\infty}^{+\infty} \frac{1}{|x|} f_1(x) f_2(\frac{z}{x}) dx$$

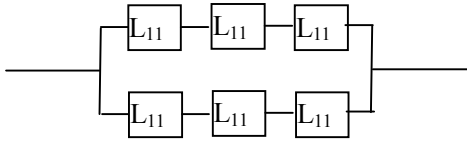
$$= \begin{cases} 0 & z < 0 \\ \int_z^1 \frac{1}{|x|} \times 1 \times 1 dx = -\ln z & 0 < z \leq 1 \\ \int_0^1 \frac{1}{|x|} \times 1 \times 0 dx = 0 & z > 1 \end{cases} = \begin{cases} -\ln z & 0 < z \leq 1 \\ 0 & \text{其他} \end{cases}$$

***极值分布** 设随机变量 X 和 Y 相互独立，分布函数分别为 $F_X(x)$ 和 $F_Y(y)$, 求 $Z=\max(X, Y)$, $Z=\min(X, Y)$ 的分布函数.

$$\text{解 } F_Z(z) = P\{\max(X, Y) \leq z\} = P\{X \leq z, Y \leq z\} = P\{\max(X, Y) \leq z\} = P\{X \leq z, Y \leq z\}$$

$$F_Z(z) = P\{\min(X, Y) \leq z\} = 1 - P\{X > z, Y > z\} \\ = 1 - P\{X > z\} P\{Y > z\} = 1 - (1 - F_X(z))(1 - F_Y(z))$$

例 9、系统如图, 每个元件寿命服从 $E(\lambda)$, 求系统寿命的概率密度.



$$\text{解 } L_{ij} \sim E(\lambda), F_{ij}(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases};$$

$$\eta_i = \min(L_{i1}, L_{i2}, L_{i3}), i = 1, 2$$

$$F_{\eta_i}(y) = 1 - (1 - F(x))^3 = \begin{cases} 1 - e^{-3\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases};$$

$$\zeta = \max\{\eta_1, \eta_2\}, F_{\zeta}(z) = F_{\eta_1}(y)F_{\eta_2}(y) = \begin{cases} (1 - e^{-3\lambda z})^2 & z > 0 \\ 0 & z \leq 0 \end{cases};$$

$$f_{\zeta}(z) = F'_{\zeta}(z) = \begin{cases} 6\lambda e^{-3\lambda z}(1 - e^{-3\lambda z}) & z > 0 \\ 0 & z \leq 0 \end{cases}.$$

第三章 大数定理与中心极限定理

第一节、计算基础

$$1、\text{ 证明 } \int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2},$$

证明 在区域 $D_R = \{(x, y) | x^2 + y^2 = R^2, x \geq 0, y \geq 0\}$ 中,

$$\iint_{D_R} e^{-(x^2+y^2)} d\sigma = \int_0^{\frac{\pi}{2}} d\theta \int_0^R e^{-r^2} r dr = \frac{\pi}{4} (1 - e^{-R^2}), \text{ 于是 } \lim_{R \rightarrow +\infty} \iint_{D_R} e^{-(x^2+y^2)} d\sigma = \frac{\pi}{4}$$

$$\text{故当 } D = [0, +\infty) \times [0, +\infty) \text{ 时, } \iint_D e^{-(x^2+y^2)} d\sigma = \frac{\pi}{4}$$

$$\text{又 } (\int_0^{+\infty} e^{-x^2} dx)^2 = \int_0^{+\infty} e^{-x^2} dx \int_0^{+\infty} e^{-y^2} dy = \iint_D e^{-(x^2+y^2)} d\sigma = \frac{\pi}{4},$$

$$\text{于是 } \int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}, \quad \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

对于 $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ (正态密度), 换元 $y = \frac{x-\mu}{\sigma\sqrt{2}}$ 得

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

$$2 \text{ 研究函数 } \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, \quad x > 0$$

解：分部积分得

$$\Gamma(x) = -t^{x-1}e^{-t} \Big|_0^{\infty} + (x-1) \int_0^{\infty} t^{x-2} e^{-t} dt = (x-1)\Gamma(x-1), \quad x > 1$$

$$\text{可得 } \Gamma(x+n) = (x+n-1)(x+n-2) \cdots x\Gamma(x)$$

$$\text{由于 } \Gamma(1) = \int_0^{\infty} e^{-t} dt = 1, \text{ 因此 } \Gamma(n) = (n-1)!\Gamma(1) = (n-1)!$$

$$\text{令 } t = x^2, \quad \Gamma(1/2) = \int_0^{\infty} t^{-1/2} e^{-t} dt = \int_0^{\infty} 2e^{-x^2} dx = \sqrt{\pi}$$

$$\Gamma(3/2) = (1/2)\Gamma(1/2) = \sqrt{\pi}/2$$

$$\Gamma(5/2) = (3/2)\Gamma(3/2) = 3\sqrt{\pi}/4$$

$$\Gamma\left(\frac{n+1}{2}\right) = \frac{n-1}{2} \cdot \frac{n-3}{2} \cdots \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{(n-1)!!}{2^{\frac{n}{2}}}$$

$$3、\text{ 设 } X \sim N(0,1) \text{ 则 } EX^n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-\frac{x^2}{2}} dx,$$

$$\text{当 } n \text{ 为奇数时, } EX^n = 0, \text{ 当 } n \text{ 为偶数时, } EX^n = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-\frac{x^2}{2}} dx$$

$$\text{令 } t = \frac{x^2}{2}, \text{ 则 } dt = x dx, x = (2t)^{\frac{1}{2}}, dx = \frac{1}{x} dt = (2t)^{-\frac{1}{2}} dt,$$

$$\begin{aligned} EX^n &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} (2t)^{\frac{n}{2}} e^{-t} (2t)^{-\frac{1}{2}} dt = \frac{2^{\frac{n}{2}}}{\sqrt{\pi}} \int_0^{+\infty} t^{\frac{n-1}{2}} e^{-t} dt \\ &= \frac{2^{\frac{n}{2}}}{\sqrt{\pi}} \int_0^{+\infty} t^{\frac{n+1}{2}-1} e^{-t} dt = \frac{2^{\frac{n}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right) = \frac{2^{\frac{n}{2}}}{\sqrt{\pi}} \times \frac{\sqrt{\pi}(n-1)!!}{2^{\frac{n}{2}}} = (n-1)!! \end{aligned}$$

$$4、\beta \text{ 函数 } \beta(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

$$\text{求证: } \beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, x > 0, y > 0$$

$$\text{证明: } \Gamma(x)\Gamma(y) = \int_0^{\infty} t^{x-1} e^{-t} dt \int_0^{\infty} u^{y-1} e^{-u} du$$

令 $u = tv$ 得

$$\begin{aligned} \Gamma(x)\Gamma(y) &= \int_0^{\infty} t^{x-1} e^{-t} dt \int_0^{\infty} t^y v^{y-1} e^{-tv} dv = \int_0^{\infty} v^{y-1} dv \int_0^{\infty} t^{x+y-1} e^{-t(1+v)} dt \\ &= \int_0^{\infty} v^{y-1} dv \int_0^{\infty} \frac{w^{x+y-1} e^{-w} dw}{(1+v)^{x+y}} = \Gamma(x+y) \int_0^{\infty} \frac{v^{y-1} dv}{(1+v)^{x+y}} \end{aligned}$$

$$\text{令 } t(1+v)=w, \text{ 则 } dt = \frac{1}{(1+v)} dw, t = \frac{w}{(1+v)}$$

$$\begin{aligned}\Gamma(x)\Gamma(y) &= \int_0^\infty v^{y-1} dv \int_0^\infty \frac{w^{x+y-1} e^{-w} dw}{(1+v)^{x+y}} \\ &= \int_0^\infty w^{x+y-1} e^{-w} dw \int_0^\infty \frac{v^{y-1} dv}{(1+v)^{x+y}} = \Gamma(x+y) \int_0^\infty \frac{v^{y-1} dv}{(1+v)^{x+y}}\end{aligned}$$

$$\text{令 } v = \tan^2 \theta, \text{ 则 } \frac{v^{y-1} dv}{(1+v)^{x+y}} = \frac{(\tan \theta)^{2y-2} 2 \tan \theta \sec^2 \theta}{(\sec \theta)^{2x+2y}} d\theta$$

$$\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = 2 \int_0^{\pi/2} (\cos \theta)^{2x-1} (\sin \theta)^{2y-1} d\theta \quad (1)$$

$$\text{令 } \lambda = \sin^2 \theta, \text{ 则 } d\lambda = 2 \sin \theta \cos \theta d\theta,$$

$$2(\cos \theta)^{2x-1} (\sin \theta)^{2y-1} d\theta = (\cos \theta)^{2x-2} (\sin \theta)^{2y-2} d\lambda = \lambda^{x-1} (1-\lambda)^{y-1} d\lambda$$

$$\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 \lambda^{x-1} (1-\lambda)^{y-1} d\lambda = \beta(x, y) \quad (3)$$

$$5、\text{ 在 } \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = 2 \int_0^{\pi/2} (\cos \theta)^{2x-1} (\sin \theta)^{2y-1} d\theta \text{ 中}$$

$$\text{设 } x=y=1/2, \text{ 得 } [\Gamma(1/2)]^2 = 2\Gamma(1) \int_0^{\pi/2} d\theta = \pi$$

$$\text{故得 } \Gamma(1/2) = \sqrt{\pi}, \text{ 由此得 } \int_0^\infty e^{-x^2} dx = \frac{1}{2} \int_0^\infty t^{-1/2} e^{-t} dt = \frac{1}{2} \sqrt{\pi} \quad (2)$$

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^\infty e^{-(x-a)^2/2\sigma} dx = 1$$

$$6、\text{ 证明沃利斯 (Wallis) 公式 } \lim_{m \rightarrow \infty} \left[\frac{(2m)!!}{(2m-1)!!} \right]^2 \frac{1}{2m+1} = \frac{\pi}{2}$$

$$\text{证明: 设 } J_n = \int_0^{\pi/2} \sin^n x dx \quad (n \in N)$$

$$\text{当 } n \geq 2 \text{ 时, } J_n = \int_0^{\frac{\pi}{2}} \sin^n x dx = -\sin^{n-1} x \cos x \Big|_0^{\frac{\pi}{2}} + (n-1) \int_0^{\frac{\pi}{2}} \sin^{n-2} x \cos^2 x dx$$

$$= (n-1) \left[\int_0^{\frac{\pi}{2}} \sin^{n-2} x dx - \int_0^{\frac{\pi}{2}} \sin^n x dx \right] = (n-1) J_{n-2} - (n-1) J_n$$

$$n J_n = (n-1) J_{n-2}, J_n = \frac{n-1}{n} J_{n-2} (n \geq 2)$$

$$\text{因 } J_1 = \int_0^{\frac{\pi}{2}} \sin x dx = 1, J_0 = \int_0^{\frac{\pi}{2}} dx = \frac{\pi}{2},$$

$$\text{故 } J_{2m} = \frac{2m-1}{2m} \cdot \frac{2m-3}{2m-2} \cdots \frac{\pi}{2} = \frac{(2m-1)!!}{(2m)!!} \cdot \frac{\pi}{2},$$

$$J_{2m+1} = \frac{2m}{2m+1} \cdot \frac{2m-2}{2m-1} \cdots \frac{2}{3} \cdot 1 = \frac{(2m)!!}{(2m+1)!!}$$

$$\text{因为 } \int_0^{\frac{\pi}{2}} \sin^{2m+1} x dx < \int_0^{\frac{\pi}{2}} \sin^{2m} x dx < \int_0^{\frac{\pi}{2}} \sin^{2m-1} x dx$$

$$\text{所以 } \frac{(2m)!!}{(2m+1)!!} < \frac{(2m-1)!!}{(2m)!!} \cdot \frac{\pi}{2} < \frac{(2m-2)!!}{(2m-1)!!}$$

$$\text{由此得 } \left[\frac{(2m)!!}{(2m+1)!!} \right]^2 \frac{1}{2m+1} < \frac{\pi}{2} < \left[\frac{(2m)!!}{(2m-1)!!} \right]^2 \frac{1}{2m}$$

$$\left[\frac{(2m)!!}{(2m-1)!!} \right]^2 \frac{1}{2m} - \left[\frac{(2m)!!}{(2m+1)!!} \right]^2 \frac{1}{2m+1} = \left[\frac{(2m)!!}{(2m-1)!!} \right]^2 \frac{1}{2m(2m+1)} < \frac{\pi}{2} \cdot \frac{1}{2m} \rightarrow 0 (m \rightarrow \infty)$$

$$\text{故得 } \lim_{m \rightarrow \infty} \left[\frac{(2m)!!}{(2m-1)!!} \right]^2 \frac{1}{2m+1} = \frac{\pi}{2}$$

$$7、\text{证明 Stirling 公式 } \Gamma(n+1) = n! \approx \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}}$$

证明: 当 $|x| < 1$ 时

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1} \frac{x^n}{n} + \cdots$$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots - \frac{x^n}{n} - \cdots$$

$$\ln \frac{1+x}{1-x} = 2x + \frac{2x^3}{3} + \frac{2x^5}{5} + \cdots = 2x \left(1 + \frac{x^2}{3} + \frac{x^4}{5} + \cdots \right)$$

$$\begin{aligned}
& \text{令 } x = \frac{1}{2n+1} \text{ 得} \\
& \ln \frac{n+1}{n} = \frac{1}{2n+1} \left[1 + \frac{1}{3} \cdot \frac{1}{(2n+1)^2} + \frac{1}{5} \cdot \frac{1}{(2n+1)^4} + \cdots \right] \\
& \left(n + \frac{1}{2}\right) \ln\left(1 + \frac{1}{n}\right) = 1 + \frac{1}{3} \cdot \frac{1}{(2n+1)^2} + \frac{1}{5} \cdot \frac{1}{(2n+1)^4} + \cdots \\
& < 1 + \frac{1}{3} \left[\frac{1}{(2n+1)^2} + \frac{1}{(2n+1)^4} + \cdots \right] = 1 + \frac{1}{12n(n+1)} \\
& 1 < \left(n + \frac{1}{2}\right) \ln\left(1 + \frac{1}{n}\right) < 1 + \frac{1}{12n(n+1)} \\
& e < \left(1 + \frac{1}{n}\right)^{n+\frac{1}{2}} < e^{1+\frac{1}{12n(n+1)}}
\end{aligned}$$

$$\text{设 } a_n = \frac{n! e^n}{n^{n+\frac{1}{2}}}, \text{ 则 } \frac{a_n}{a_{n+1}} = \frac{\left(1 + \frac{1}{n}\right)^{n+\frac{1}{2}}}{e} \in (1, e^{\frac{1}{12n(n+1)}})$$

于是 $\{a_n\}$ 递减, 有下界, 故 $\lim_{n \rightarrow \infty} a_n$ 存在, 设为 a

$$\text{由 } \frac{a_n}{a_{n+1}} < e^{\frac{1}{12n(n+1)}} = \frac{e^{\frac{1}{12n}}}{e^{\frac{1}{12(n+1)}}} \text{ 得 } \frac{a_n}{e^{\frac{1}{12n}}} < \frac{a_{n+1}}{e^{\frac{1}{12(n+1)}}}, \text{ 因当 } n \rightarrow \infty \text{ 时}$$

两边的极限都是 a , 故 $\frac{a_n}{e^{\frac{1}{12n}}} < a < a_n$,

于是存在 $\theta \in (0, 1)$, $a_n = a e^{\frac{\theta}{12n}}$ (θ 随 n 的变化而变化)

$$\text{故 } \frac{n! e^n}{n^{n+\frac{1}{2}}} = a e^{\frac{\theta}{12n}}, n! = \frac{a n^{n+\frac{1}{2}} e^{\frac{\theta}{12n}}}{e^n} = a \sqrt{n} \left(\frac{n}{e}\right)^n e^{\frac{\theta}{12n}} = \sqrt{n} \left(\frac{n}{e}\right)^n a_n$$

由 Wallis 公式得

$$\begin{aligned}
\sqrt{\frac{\pi}{2}} &= \lim_{m \rightarrow \infty} \frac{1}{\sqrt{2n+1}} \frac{(2n)!!}{(2n-1)!!} = \lim_{m \rightarrow \infty} \frac{1}{\sqrt{2n+1}} \frac{[(2n)!!]^2}{(2n)!} \\
&= \lim_{m \rightarrow \infty} \frac{1}{\sqrt{2n+1}} \frac{2^{2n} (n!)^2}{(2n)!} = \lim_{m \rightarrow \infty} \frac{1}{\sqrt{2n+1}} \frac{2^{2n} \left[\sqrt{n} \left(\frac{n}{e}\right)^n a_n\right]^2}{\sqrt{2n} \left(\frac{2n}{e}\right)^{2n} a_{2n}} \\
&= \lim_{m \rightarrow \infty} \frac{\sqrt{n}}{\sqrt{4n+2}} \frac{a_n^2}{a_{2n}} = \frac{a}{2}, \text{ 于是 } a = \sqrt{2\pi} \\
&\text{因此 } n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{\theta}{12n}} \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n
\end{aligned}$$

例 抛 100 个分币, 求有 50 个正面的概率。

$$\text{解 } C_{2n}^n \frac{1}{2^{2n}} = \frac{(2n)!}{n! n!} \frac{1}{2^{2n}} \approx \frac{(2n/e)^{2n} \sqrt{4\pi n}}{(n/e)^{2n} 2\pi n} \frac{1}{2^{2n}} = \frac{1}{\sqrt{\pi n}}$$

$$C_{100}^{50} \frac{1}{2^{100}} \approx \frac{1}{\sqrt{50\pi}} = 0.08$$

8、特征函数

(1)定义 设 ξ 为实随机变量,称 $f(t)=Ee^{it\xi}$ 为 ξ 的特征函数这里 t 是任意实数.

由于 $E|e^{it\xi}|=1$, 因此对任意 ξ , 对一切 $t \in (-\infty, \infty)$, (1)式都有意义. 换句话说, 对每个随机变量 ξ (或者说每个分布函数 $F(x)$), 都有一个特征函数 $f(t)$ 与之对应, 它是定义在 $(-\infty, \infty)$ 上的实变量复值函数.

特征函数是 ξ 的函数 $e^{it\xi}$ 的数学期望, $f(t)=\int_{-\infty}^{\infty} e^{itx} dF(x)$.

特别, 若 ξ 为离散型, $P(\xi=x_n)=p_n$, $n=1,2,\dots$, 则 $f(t)=\sum_{n=1}^{\infty} p_n e^{itx_n}$.

若 ξ 是连续型, 其密度为 $p(x)$, 则 $f(t)=\int_{-\infty}^{\infty} e^{itx} p(x)dx$, 它就是函数 $p(x)$ 的傅里叶变换.

(2)五个重要分布的特征函数

第一、退化分布 $P(\xi=c)=1$ 的特征函数 $f(t)=e^{ict}$.

第二、二项分布 $B(n, p)$ 的特征函数

$$f(t)=\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{itk} = \sum_{k=0}^n \binom{n}{k} (pe^{it})^k q^{n-k} = (pe^{it} + q)^n, \quad (p+q=1).$$

第三、泊松分布 $P(\lambda)$ 的特征函数 $f(t)=\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} e^{itk} = \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} e^{-\lambda} = e^{\lambda(e^{it}-1)}$.

第四、均匀分布 $U[a, b]$ 的特征函数 $f(t) = \int_a^b \frac{1}{b-a} e^{itx} dx = \frac{e^{itb} - e^{ita}}{(b-a)it}$.

第五、正态分布 $N(a, \sigma^2)$ 的特征函数 $f(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{itx - (x-a)^2/2\sigma^2} dx = e^{iat - \sigma^2 t^2/2}$.

第二节 大数定理与中心极限定理

一、大数定理

1、切比雪夫不等式

设随机变量 X 的数学期望 $E(X)$ 与方差 $D(X)$ 存在, 则对于任意正数 ε ,

$$P[|X - E(X)| \geq \varepsilon] \leq \frac{D(X)}{\varepsilon^2} \quad \text{或} \quad P[|X - E(X)| < \varepsilon] \geq 1 - \frac{D(X)}{\varepsilon^2}$$

证明: (1) 设 X 为离散型随机变量, 其分布列: $P(X=X_k)=P_k (k=1,2,\dots)$ 则

$$P[|X - E(X)| \geq \varepsilon] = \sum_{|X_k - E(X)| \geq \varepsilon} P_k \leq \sum_{|X_k - E(X)| \geq \varepsilon} \frac{(X_k - E(X))^2}{\varepsilon^2} P_k$$

$$\leq \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} (X_k - E(X))^2 P_k = \frac{D(X)}{\varepsilon^2}$$

(2) 设 $f(x)$ 为 X 的密度函数, 记 $E(X) = \mu$, $D(X) = \sigma^2$ 则

$$P\{|X - E(X)| \geq \varepsilon\} = \int_{|x-\mu| \geq \varepsilon} f(x) dx \leq \int_{|x-\mu| \geq \varepsilon} \frac{(x-\mu)^2}{\varepsilon^2} f(x) dx$$

$$\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \leq \frac{1}{\varepsilon^2} \times \sigma^2 = \frac{D(X)}{\varepsilon^2}$$

例 1、投掷一枚硬币, 为了至少有 90% 的把握使正面向上的频率在 0.49 与 0.51 之间, 试估计需要的投掷次数 n 。

解: 用 X 表示在 n 次实验中出现正面的次数, 显然 $X \sim B(n, 0.5)$ 。那么

$$E(X) = np = 0.5n$$

$$D(X) = np(1-p) = 0.25n$$

n 次试验中事件 A 出现的频率为 $f_n = \frac{X}{n}$

由切比雪夫不等式得

$$P(0.49 < \frac{X}{n} < 0.51) = P(|X - 0.5n| < 0.01n) \geq 1 - \frac{0.25n}{(0.01n)^2} = 1 - \frac{2500}{n}$$

由题意可知 $1 - \frac{2500}{n} \geq 0.9$ 解得 $n \geq 25000$

即至少要投掷这枚硬币 25000 次, 才能至少有 90% 的把握使正面向上的频率在 0.49 与 0.51 之间。

2、切比雪夫大数定律

设独立随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 的数学期望

$E(X_1), E(X_2), \dots, E(X_n), \dots$ 和方差 $D(X_1), D(X_2), \dots, D(X_n), \dots$ 都存在, 并且方差

是一致有上界的, 即存在常数 C , 使得 $D(X_i) \leq C, i = 1, 2, \dots, n, \dots$

则对于任意的正数 ε , 有 $\lim_{n \rightarrow \infty} P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) = 1$ 。

证明: $E(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{1}{n} \sum_{i=1}^n E(X_i)$

因 $X_1, X_2, \dots, X_n$ 相互独立, 故 $D(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{1}{n^2} \sum_{i=1}^n D(X_i)$

由切比雪夫不等式得 $P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) \geq 1 - \frac{1}{n^2 \varepsilon^2} \sum_{i=1}^n D(X_i)$

因 $D(X_i) \leq C (i=1, 2, \dots, n)$, 故 $\sum_{i=1}^n D(X_i) \leq nC$, 由此得

$$P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) \geq 1 - \frac{C}{n\varepsilon^2}$$

故 $\lim_{n \rightarrow \infty} P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) \geq 1$, 又 $\lim_{n \rightarrow \infty} P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) \leq 1$ 。

于是 $\lim_{n \rightarrow \infty} P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) = 1$

3、伯努利大数定理

设 u_n 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 在每次试验中发生的概率, 则

对于任意正整数 ε , 有 $\lim_{n \rightarrow \infty} P\left\{\left|\frac{u_n}{n} - p\right| < \varepsilon\right\} = 1$.

证明: 令 $X_k = \begin{cases} 1 & \text{第 } k \text{ 次试验 } A \text{ 发生} \\ 0 & \text{第 } k \text{ 次试验 } A \text{ 不发生} \end{cases} (k=1, 2, \dots)$, $X_1, X_2, \dots, X_n, \dots$ 是 n 个相互独立的

的随机变量, 且 $E(X_i) = p, D(X_i) = pq = p(1-p) \leq \frac{1}{4}$.

于是对于任意正整数 ε , 有 $\lim_{n \rightarrow \infty} P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) = 1$

因 $\frac{1}{n} \sum_{i=1}^n X_i = \frac{u_n}{n}$, $\frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \times np = p$, 故 $\lim_{n \rightarrow \infty} P\left\{\left|\frac{u_n}{n} - p\right| < \varepsilon\right\} = 1$

4、普阿松大数定理 如果在一个独立试验序列中, 事件 A 在第 k 次试验中出现的概率等于 p_k , 以 u_k 记在前 n 次试验中事件 A 出现的次数, 则对任意正整数 ε , 有

$$\lim_{n \rightarrow \infty} P(|\frac{u_k}{n} - \frac{1}{n} \sum_{i=1}^n p_k| < \varepsilon) = 1$$

证明: 设第 k 次试验事件 A 发生的次数为 X_k , $E(X_i) = p_k < 1$, $D(X_i) = p_k(1-p_k) \leq \frac{1}{4}$

于是对于任意正整数 ε , 有 $\lim_{n \rightarrow \infty} P(|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)| < \varepsilon) = 1$

因 $\frac{1}{n} \sum_{i=1}^n X_i = \frac{u_k}{n}$, $\frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n p_k$ 于是 $\lim_{n \rightarrow \infty} P(|\frac{u_k}{n} - \frac{1}{n} \sum_{i=1}^n p_k| < \varepsilon) = 1$

5、两个概念

(1) 设 $Y_1, Y_2, Y_3, \dots, Y_n, \dots$ 是一随机变量序列, a 为一常数, 对于任意正整数 ε , 有

$\lim_{n \rightarrow \infty} P(|Y_n - \frac{1}{n} \sum_{i=1}^n p_k| < \varepsilon) = 1$ 则称随机变量序列 $Y_1, Y_2, Y_3, \dots, Y_n, \dots$ 依概率收敛于 a

记作 $Y_n \xrightarrow{P} a$

(2) 设 $X_1, X_2, X_3, \dots, X_n, \dots$ 是一随机变量序列, a 为一常数, 对于任意正整数 ε , 有

$Y_n = \frac{1}{n} \sum_{i=1}^n X_k$, 若存在常数 a_k , 使对于任意正整数 ε , 有 $\lim_{n \rightarrow \infty} P(|Y_n - a_k| < \varepsilon) = 1$ 则称随机

变量序列 $X_1, X_2, X_3, \dots, X_n, \dots$ 服从大数定理

二、中心极限定理

1、德莫佛—拉普拉斯定理

设 u_n 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 在每次试验中发生的概率, 则

对任意的有限区间 $[a, b]$.

(1) 当 $a \leq \frac{k-np}{\sqrt{npq}} \leq b$ 时, 令 $x_k = \frac{k-np}{\sqrt{npq}} (q=1-p)$, 则当 $n \rightarrow \infty$ 时有

$$P(\mu_n = k) \div (\frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2}) \rightarrow 1$$

(2) 由(1)当 $n \rightarrow \infty$ 时, 一致地有 $P(a \leq \frac{\mu_n - np}{\sqrt{npq}} < b) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$

证明:(1) 因为 $m! = \sqrt{2\pi m} m^m e^{-m} e^{\theta_m} (0 < \theta_m < \frac{1}{12m})$, $x_k = \frac{k-np}{\sqrt{npq}} (q=1-p)$

则 $k = np + x_k \sqrt{npq}$, 取 $j = n - k = k = np - x_k \sqrt{npq}$

$$\begin{aligned} \text{故 } P(\mu_n = k) &= \frac{n!}{k!j!} p^k q^j = \frac{\sqrt{2\pi n} n^n e^{-n} e^{\theta_n}}{\sqrt{2\pi k} k^k e^{-k} e^{\theta_k} \sqrt{2\pi j} j^j e^{-j} e^{\theta_j}} p^k q^j \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{kj}} \left(\frac{np}{k}\right)^k \left(\frac{np}{j}\right)^j e^{\theta} \end{aligned}$$

$$\text{这里 } 0 < \theta_n < \frac{1}{12n}, 0 < \theta_k < \frac{1}{12k}, 0 < \theta_j < \frac{1}{12j}, |\theta| = |\theta_n - \theta_k - \theta_j| < \frac{1}{12} \left(\frac{1}{n} + \frac{1}{k} + \frac{1}{j}\right)$$

$$\begin{aligned} \sqrt{\frac{n}{kj}} &= \left(\frac{kj}{n}\right)^{-\frac{1}{2}} = \left[\frac{(np + x_k \sqrt{npq})(nq - x_k \sqrt{npq})}{n}\right]^{-\frac{1}{2}} = \left[\frac{np(1 + x_k \sqrt{\frac{q}{np}})nq(1 - x_k \sqrt{\frac{p}{np}})}{n}\right]^{-\frac{1}{2}} \\ &= \frac{1}{\sqrt{npq}} \left[(1 + x_k \sqrt{\frac{q}{np}})(1 - x_k \sqrt{\frac{p}{np}})\right]^{-\frac{1}{2}} = \frac{1}{\sqrt{npq}} \left[1 + \frac{x_k(q-p)}{\sqrt{npq}} + o\left(\frac{1}{n}\right)\right]^{-\frac{1}{2}} \\ &= \frac{1}{\sqrt{npq}} \left[1 - \frac{x_k(q-p)}{2\sqrt{npq}} + o\left(\frac{1}{n}\right)\right] \end{aligned}$$

由 $k = np + x_k \sqrt{npq}$, $j = n - k = nq - x_k \sqrt{npq}$ 得

$$\frac{k}{np} = 1 + x_k \sqrt{\frac{q}{np}}, \frac{j}{nq} = 1 - x_k \sqrt{\frac{p}{nq}},$$

于是

$$\begin{aligned} \ln\left[\left(\frac{np}{k}\right)^k \left(\frac{np}{j}\right)^j\right] &= -k \ln \frac{k}{np} - j \ln \frac{j}{nq} \\ &= -(np + x_k \sqrt{npq}) \left[x_k \sqrt{\frac{q}{np}} - \frac{qx_k^2}{2np} + \frac{1}{3} \left(\frac{q}{np}\right)^{\frac{3}{2}} x_k^3 + o\left(\frac{1}{n^2}\right)\right] \\ &\quad + (nq - x_k \sqrt{npq}) \left[x_k \sqrt{\frac{p}{nq}} - \frac{px_k^2}{2nq} + \frac{1}{3} \left(\frac{p}{nq}\right)^{\frac{3}{2}} x_k^3 + o\left(\frac{1}{n^2}\right)\right] \\ &= -\frac{1}{2} x_k^2 + \frac{q-p}{6\sqrt{npq}} x_k^3 + o\left(\frac{1}{n}\right) \\ \left(\frac{np}{k}\right)^k \left(\frac{np}{j}\right)^j &= e^{-\frac{1}{2} x_k^2} e^{\frac{q-p}{6\sqrt{npq}} x_k^3} e^{o\left(\frac{1}{n}\right)} = e^{-\frac{1}{2} x_k^2} \left[1 + \frac{q-p}{6\sqrt{npq}} x_k^3 + o\left(\frac{1}{n}\right)\right] \left[1 + o\left(\frac{1}{n}\right)\right] \end{aligned}$$

$$\begin{aligned} P(\mu_n = k) &= \frac{1}{\sqrt{2\pi k}} \sqrt{\frac{n}{kj}} \left(\frac{np}{k}\right)^k \left(\frac{np}{j}\right)^j e^{\theta} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{1}{2} x_k^2} \left[1 - \frac{x_k(q-p)}{2\sqrt{npq}} + o\left(\frac{1}{n}\right)\right] \left[1 + \frac{q-p}{6\sqrt{npq}} x_k^3 + o\left(\frac{1}{n}\right)\right] \left[1 + o\left(\frac{1}{n}\right)\right] \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{1}{2} x_k^2} \left[1 + o\left(\frac{1}{\sqrt{n}}\right)\right] \end{aligned}$$

$$\text{于是当 } n \rightarrow \infty \text{ 时有 } P(\mu_n = k) \div \left(\frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x_k^2}\right) \rightarrow 1$$

此式有下面两种表达

$$\text{①对任意的正数 } \varepsilon, \text{ 有 } P(\mu_n = k) = \frac{1}{\sqrt{npq}} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x_k^2} + \varepsilon\right)$$

$$\textcircled{2} \lim_{n \rightarrow \infty} C_n^k p^k q^{n-k} \div \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{k-np}{\sqrt{npq}})^2} = 1$$

$$(2) \text{记 } k_1 = \min\{\mu_n \mid a \leq \frac{\mu_n - np}{\sqrt{npq}} < b\}, k_2 = \max\{\mu_n \mid a \leq \frac{\mu_n - np}{\sqrt{npq}} < b\},$$

$$P(a \leq \frac{\mu_n - np}{\sqrt{npq}} < b) = \frac{1}{\sqrt{npq}} \sum_{k=k_1}^{k_2} (\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2} + \varepsilon) = \frac{1}{\sqrt{npq}} \sum_{k=k_1}^{k_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_k^2} + \frac{\varepsilon(k_2 - k_1 + 1)}{\sqrt{npq}}$$

$$\text{因 } 0 < \frac{\varepsilon(k_2 - k_1 + 1)}{\sqrt{npq}} < \frac{(b-a)\sqrt{npq} + 1}{\sqrt{npq}} \varepsilon = [(b-a) + \frac{1}{\sqrt{npq}}] \varepsilon$$

$$\text{于是当(1)当 } n \rightarrow \infty \text{ 时,一致地有 } \lim_{n \rightarrow \infty} P(a \leq \frac{\mu_n - np}{\sqrt{npq}} < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

$$\text{一般地 } \lim_{n \rightarrow \infty} P(\frac{\mu_n - np}{\sqrt{npq}} < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

$$\text{注(1)设 } F_n(x) = P(\frac{\mu_n - np}{\sqrt{npq}} < x), \text{ 则 } \lim_{n \rightarrow \infty} F_n(x) = \Phi(x)$$

$$(2) P(k_1 \leq \mu_n \leq k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k} \approx \Phi(\frac{k_2 - np}{\sqrt{npq}}) - \Phi(\frac{k_1 - np}{\sqrt{npq}})$$

$$(3) \text{ 当 } z > 0 \text{ 时, } \int_z^{+\infty} e^{-\frac{x^2}{2}} dx < \frac{1}{z} \int_z^{+\infty} x e^{-\frac{x^2}{2}} dx = \frac{1}{z} e^{-\frac{z^2}{2}}, \text{ 当 } z=10 \text{ 时就很小了}$$

例 1、某单位内部有 260 部电话分机，每个分机有 4%的时间要与外线通话，可以认为每个电话分机用不同的外线是相互独立的，问总机需备多少条外线才能 95%满足每个分机在用外线时不用等候？

解 令 $X_k = \begin{cases} 1 & \text{第 } k \text{ 个分机要用外线} \\ 0 & \text{第 } k \text{ 个分机不要用外线} \end{cases} (k=1, 2, \dots, 260)$, $X_1, X_2, \dots, X_{260}$ 是 260 个相

互独立的随机变量，且 $E(X_i) = 0.04$, $m = X_1 + X_2 + \dots + X_{260}$ 表示同时使用外线的分机数，

根据题意应确定最小的 x 使 $P\{m < x\} \geq 95\%$ 成立。由上面定理，有

$$P\{m < x\} = P\left\{\frac{m - 260p}{\sqrt{260p(1-p)}} \leq \frac{x - 260p}{\sqrt{260p(1-p)}}\right\} \approx \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

查得 $\Phi(1.65) = 0.9505 > 0.95$ ，故，取 $b = 1.65$ ，于是

$$x = b\sqrt{260p(1-p)} + 260p = 1.65 \times \sqrt{260 \times 0.04 \times 0.96} + 260 \times 0.04 \approx 15.61$$

也就是说，至少需要 16 条外线才能 95%满足每个分机在用外线时不用等候。

例 2、某车间有 200 台车床，它们独立地工作着，开工率各为 0.6，开工时耗电各为 1 千瓦，问供电所至少要给这个车间多少电力才能以 99.9%的概率保证这个车间不会因供电不足而影响生产？

解:设有 ξ 台正常中,则 $\xi \sim B(200, 0.6)$, 要求 r 使

$$P(\xi \leq r) = \sum_{k=0}^r \binom{200}{k} 0.6^k 0.4^{200-k} \geq 0.999$$

$$\phi\left(\frac{r-200 \times 0.6}{\sqrt{200 \times 0.6 \times 0.4}}\right) - \phi\left(\frac{-200 \times 0.6}{\sqrt{200 \times 0.6 \times 0.4}}\right)$$

于是 $= \phi\left(\frac{r-120}{\sqrt{48}}\right) - \phi(-17.32) \approx \phi\left(\frac{r-120}{\sqrt{48}}\right) \geq 0.99$

$$\text{查表得 } \frac{r-120}{\sqrt{48}} = 3.1, r = 141$$

于是供电 141 千瓦由于供电不足影响生产的可能性小于 0.001

例 3 用机器包装味精, 每袋净重为随机变量, 期望值为 100 克, 标准差为 10 克, 一箱内装 200 袋味精, 求一箱味精净重大于 20500 克的概率。

解 设一箱味精净重为 X 克, 箱中第 k 袋味精的净重为 X_k 克, $k = 1, 2, \dots, 200$.

$X_1, X_2, \dots, X_{200}$ 是 200 个相互独立的随机变量, 且 $E(X_k) = 100, D(X_k) = 100$,

$$E(X) = E(X_1 + X_2 + \dots + X_{200}) = 20000, D(X) = 20000, \sqrt{D(X)} = 100\sqrt{2}$$

因而有 $P(\{X > 20500\}) = 1 - P(\{X < 20500\})$

$$= 1 - P\left\{\frac{X - 20000}{100\sqrt{2}} \leq \frac{500}{100\sqrt{2}}\right\} \approx 1 - \Phi(3.54) = 0.0002$$

2、辛钦定理

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 服从同一分布, 且具有数学期望

$$E(X_k) = \mu \quad (k = 1, 2, \dots), \text{ 则对于任意正数 } \varepsilon, \text{ 有 } \lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| < \varepsilon\right\} = 1.$$

伯努利定理是辛钦定理的特殊情况. 辛钦定理不要求方差

证明: 由于随机变量 $X_1, X_2, \dots, X_n$ 相互独立, 服从同一分布, 故有同一特征函数, 设

为 $f(t)$, 因为数学期望存在, 故 $f(t)$ 可展开成 $f(t) = f(0) + f'(0)t + o(t) = 1 + iat + o(t)$

$$\text{而 } \frac{1}{n} \sum_{k=1}^n X_k \text{ 的特征函数为 } [f(\frac{t}{n})]^n = [1 + iat + o(t)]^n \xrightarrow{n \rightarrow \infty} e^{iat}$$

函数 e^{iat} 是连续函数, 它是退化分布 $I(X-a)$ 所对应的特征函数, 由逆极限定理知 $\frac{1}{n} \sum_{k=1}^n X_k$

的分布函数弱收敛于常数 a , 于是定理成立

例 1 (用蒙特卡洛方法计算定积分) $J = \int_a^b g(x)dx$

解: 任取一列相互独立的, 都具有在 $[a, b]$ 中均匀分布的随机变量 $\{\xi_i\}$, 则 $\{g(\xi_i)\}$

也是一列相互独立相同分布的随机变量, 而且 $Eg(\xi_i) = \frac{1}{b-a} \int_a^b g(x)dx = \frac{J}{b-a}$

$J = (b-a)Eg(\xi_i)$, 因此只要能求得 $Eg(\xi_i)$, 便能得到 J 的数值. 为了求 $Eg(\xi_i)$, 可用大数定

律, $\frac{g(\xi_1) + g(\xi_2) + \cdots + g(\xi_n)}{n} \xrightarrow{P} Eg(\xi_i)$

3、林德贝格—勒维中心极限定理

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 服从同一分布, 且具有数学期望和方差: $E(X_k) = \mu, D(X_k) = \sigma^2 > 0 (k=1, 2, \dots)$, 则随机变量之和的标准化变量

$Y_n = \frac{\sum_{k=1}^n X_k - E\left(\sum_{k=1}^n X_k\right)}{\sqrt{D\left(\sum_{k=1}^n X_k\right)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}$ 的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P\{Y_n \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

证明: 设 $X_k - \mu$ 的特征函数为 $\varphi(t)$, 则

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} = \sum_{k=1}^n \frac{X_k - \mu}{\sigma\sqrt{n}} \text{ 的特征函数为 } \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n$$

又因为 $E(X_k - \mu) = 0, D(X_k - \mu) = \sigma^2$, 所以 $\varphi'(0) = 0, \varphi''(0) = -\sigma^2$

于是特征函数 $\varphi(t)$ 有展开式

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \varphi''(0)\frac{t^2}{2} + o(t^2) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2).$$

从而对任意的 t 有,

$$\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right]^n \rightarrow e^{-\frac{t^2}{2}}, n \rightarrow \infty.$$

而 $e^{-\frac{t^2}{2}}$ 是 $N(0,1)$ 分布的特征函数, 由连续定理可知

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

成立，证毕.