高中统计学·第贰课练习

何濯羽

2024 年 2 月 14 日

1 离散型随机变量的 CDF

随机变量 X 代表某学校任意一个寝室里生病的学生人数。X 的概率分布如下方概率分布列所示。

X 的取值 x	0	1	2	3	4
$Pr(X=x)$	0.35	0.35	0.15	a	0.05

1) a 的值是多少?
2) 请计算 $Pr(X>2)$、$Pr(X\leq 3)$、$Pr(1<X<4)$。
3) 请写出 X 的累积概率函数（定义域为 $\mathbb{R}$）。

2 PMF 和 CDF

随机变量 X 代表某工厂 12 小时内生产的次品数量。X 的累积分布函数解析式为

$$F_X(x)=\begin{cases}0 & (x<0)\\ 0.25 & (0\leq x<1)\\ 0.6 & (1\leq x<2)\\ 0.95 & (2\leq x<3)\\ 1 & (x\geq 3)\end{cases}$$

请写出 X 的概率质量函数（定义域为支撑集 $\mathcal{X}$）的解析式和概率分布列。

3 分位数的计算

甲市在 2024 年 1 月上旬和中旬每天的空气质量指数如下：

35	54	80	86	72	10	66	46	36	18
88	54	79	14	16	40	59	67	35	62

请计算该数据的第 3 样本四分位数。

4 CDF 的非递减性【拓展】

求证：累积分布函数 $F(\cdot)$ 总是非单调递减的。

答案

离散型随机变量

（1）根据 PMF 的性质，我们必须有

$$\sum_{k=0}^{4} Pr(X=k) = 0.35 + 0.35 + 0.15 + a + 0.05 = 1$$

因此，$a = 0.1$。

（2）

$$\begin{aligned} Pr(X > 2) &= Pr(X=3) + Pr(X=4) = 0.15 \\ Pr(X \le 3) &= 1 - Pr(X=4) = 0.95 \\ Pr(1 < X < 4) &= Pr(X=2) + Pr(X=3) = 0.25 \end{aligned}$$

（3）

$$F_X(x) = \begin{cases} 0 & (x < 0) \\ 0.35 & (0 \le x < 1) \\ 0.7 & (1 \le x < 2) \\ 0.85 & (2 \le x < 3) \\ 0.95 & (3 \le x < 4) \\ 1 & (x \ge 4) \end{cases}$$

PMF 和 CDF

$$f_X(x) = \begin{cases} 0.25 & (x = 0) \\ 0.35 & (x = 1) \\ 0.35 & (x = 2) \\ 0.05 & (x = 3) \end{cases}$$

X 的取值 x	0	1	2	3
$Pr(X=x)$	0.25	0.35	0.35	0.05

分位数的计算

我们先对数据进行从小到大的排序：

10	14	16	18	35	35	36	40	46	54
54	59	62	66	67	72	79	80	86	88

我们有

$$p\% = 75\%$$
$$n = 20$$
$$i = n \times p\% = 15 \in \mathbb{Z}$$

因为 i 是整数，所以第 3 样本四分位数为第 15 个观测值和第 16 个观测值的平均数：

$$Q(0.75) = \frac{67 + 72}{2} = 69.5$$

CDF 的非递减性

假设 X 是一个以 $\mathcal{X}$ 为支撑集的随机变量。从实数集中任取两个实数，x_1 和 x_2，满足 $x_1 < x_2$。令

$$A = \{x \in \mathcal{X} : x \leq x_1\}$$
$$B = \{x \in \mathcal{X} : x \leq x_2\}$$

显然，$A \subseteq B$。根据概率三大公理的推论（见第壹课幻灯片），$Pr(A) \leq Pr(B)$，即

$$Pr(X \leq x_1) \leq Pr(X \leq x_2)$$

根据 CDF 的定义，我们有

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

故 CDF 是非单调递减的。