

3 离散型随机变量

3.1 离散型随机变量及分布列

对于很多实际的问题, 可以将试验结果用数值表示, 有些试验结果本身就是数值, 例如:

- 抛一枚骰子的点数 $\{1, 2, 3, \dots, 6\}$;
- 一年国家出生的婴儿数 $\{1, 2, \dots, n, \dots\}$.

有些试验结果可能与数值无关, 但可以将结果数值化, 例如:

- 抛一枚硬币, 正面朝上用0表示, 正面朝下用1表示;
- 流星坠落地球的落脚点用坐标纬度表示.

当试验结果用数值表示, 可以引入一个变量来表示随机事件, 由此产生随机变量的概念.

将样本空间 Ω 中每个样本点 ω 与实数 $X(\omega)$ 相互对应, $X(\omega)$ 是 ω 的实值函数. 称实值函数 $X(\omega) : \Omega \rightarrow \mathbb{R}$ 为随机变量(random variable), 简写为r.v., 一般用大写字母 X, Y, Z 表示. $X(\omega)$ 随样本点 ω 的不同而取不同的值, 例如:

- 抛一枚骰子, 令 $X =$ 出现的点数, 则 X 是随机变量, X 的取值为 $1, 2, \dots, 6$, 集合 $\{X \leq 4\}$ 表示点数不超过4的事件; $\{X \text{ 为偶数}\}$ 表示出现偶数点的事件.
- 用 X 表示一盏电灯的寿命, 则 X 是随机变量, 取值为 $[0, +\infty)$. 集合 $\{X \leq 500\}$ 表示电灯寿命不超过500的事件.

引入随机变量, 可以通过数学形式化后的随机变量来描述随机现象或随机事件, 利用更多数学工具. 例如: $\{X \leq -\infty\}$ 表示不可能事件; $\{X \leq +\infty\}$ 表示必然事件.

随机变量的分类: 离散型随机变量和连续型随机变量,

- **离散型随机变量:** 随机变量取值是有限的, 或无限可列的.
- **非离散型随机变量:** 随机变量取值无限且不可列的.

定义3.1. 若随机变量 X 的取值是有限个或无限可列个, 称 X 为离散型随机变量. 设随机变量 X 的取值为 $x_1, x_2, \dots, x_n, \dots$, 考虑 X 每个取值的概率为 $p_k = P(X = x_k)$ ($k = 1, 2, \dots$), 称之为 X 的分布列, 常用表格表示为

X	x_1	x_2	\dots	x_n	\dots
P	p_1	p_2	\dots	p_n	\dots

根据概率的定义可得

性质3.1. 对于随机变量 X 的分布列 $p_k = P(X = x_k)$, 有 $p_k \geq 0$ ($k \geq 1$) 以及 $\sum_k p_k = 1$.

例3.1. 设随机变量 X 的分布列为 $P(X = k) = c \left(\frac{1}{4}\right)^k$ ($k = 0, 1, 2, \dots$), 求常数 c .

解. 根据分布列的性质有

$$1 = \sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} c \left(\frac{1}{4}\right)^k = c \frac{1}{1 - 1/4} = \frac{4}{3}c$$

从而得到 $c = 3/4$. □

例3.2. 从 $\{1, 2, \dots, 10\}$ 中任取5个, 令随机变量 X 表示取出5个数中的最大值, 求 X 的分布列.

解. 由题意可知 X 的取值为5, 6, 7, 8, 9, 10, 且 $P(X = k) = \binom{k-1}{4} / \binom{10}{5}$ ($5 \leq k \leq 10$). □

X	5	6	7	8	9	10
P	$\frac{1}{252}$	$\frac{5}{252}$	$\frac{15}{252}$	$\frac{35}{252}$	$\frac{70}{252}$	$\frac{126}{252}$

例3.3. 给定常数 $\lambda > 0$, 随机变量 X 的分布列 $p_i = c\lambda^i/i!$ ($i \geq 0$), 求 $P(X > 2)$.

解. 根据分布列的性质有

$$1 = \sum_{i=0}^{\infty} p_i = c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = c \cdot e^\lambda$$

从而得到 $c = e^{-\lambda}$, 进一步得到

$$P(X > 2) = 1 - P(X \leq 2) = 1 - p_0 - p_1 - p_2 = 1 - e^{-\lambda}(1 + \lambda + \lambda^2/2).$$

□

3.2 离散型随机变量的期望和方差

给定一随机变量 X , 由于 X 的取值具有一定的随机性, 在研究中希望研究 X 的一些本质特征, 即刻画随机变量特征的不变量, 两种最常见的特征是期望与方差.

3.2.1 期望

定义3.2. 设离散型随机变量 X 的分布律为 $P(X = x_k) = p_k$ ($k = 1, 2, \dots$). 若级数 $\sum_{k=1}^{\infty} x_k \cdot p_k$ 绝对收敛, 称级数和为随机变量 X 的期望(expectation), 记为 $E(X)$, 即

$$E(X) = \sum_{k=1}^{\infty} x_k \cdot p_k.$$

期望又被称为均值(mean), 或加权平均(weighted average).

期望反映随机变量 X 所取的平均值, 值得注意的是:

- 期望 $E(X)$ 完全由 X 的概率分布决定, 不是变量, 其本质是随机变量的值 x_i 根据概率 p_i 加权所得.
- 级数的绝对收敛保证了级数和不随级数各项次序的改变而改变, 期望 $E(X)$ 反映 X 可能值的平均值, 不应随次序改变而改变.

根据定义求期望的步骤: i) 求随机变量 X 的分布律; ii) 根据定义求期望.

例3.4. 随机掷一枚骰子, X 表示点数, 其取值为 $\{1, 2, 3, 4, 5, 6\}$, 且每点等可能发生, 其分布列为 $P(X = i) = 1/6$ ($i \in [6]$). 因此随机变量 X 的期望 $E(X) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$.

例3.5. 有4个盒子编号分别为1, 2, 3, 4. 将3个不同的球随机放入4个盒子中, 用 X 表示有球盒子的最小号码, 求 $E(X)$.

解. 先求 X 的分布律

$$\begin{aligned} P(X=1) &= \frac{\binom{3}{1}3^2 + \binom{3}{2}3 + 1}{4^3} = \frac{37}{64}, & P(X=2) &= \frac{\binom{3}{1}2^2 + \binom{3}{2}2 + 1}{4^3} = \frac{19}{64}, \\ P(X=3) &= \frac{\binom{3}{1} + \binom{3}{2} + 1}{4^3} = \frac{7}{64}, & P(X=4) &= \frac{1}{64}. \end{aligned}$$

进一步可得

$$E(X) = \frac{37}{64} + 2 \cdot \frac{19}{64} + 3 \cdot \frac{7}{64} + 4 \cdot \frac{1}{64} = \frac{25}{16}.$$

□

例3.6. 一串钥匙有 n 把, 只有一把能打开门, 随机选取一把试开门, 若打不开则除去, 求打开门次数的平均数.

解. 设随机变量 X 表示试开门的次数, 则可得其分布列

$$P(X=k) = \frac{\binom{n-1}{k-1}}{\binom{n}{k-1}} \cdot \frac{1}{n-k+1} = \frac{1}{n}$$

于是有打开门次数的平均数 $E(X) = \sum_{k=1}^n k/n = (1+n)n/(2n) = (n+1)/2$.

□

根据期望的定义, 可以得到如下性质:

性质3.2. 若随机变量 $X \equiv c$, 则 $E(c) = c$; 对常数 a, b 和随机变量 X , 有 $E(aX + b) = aE(X) + b$.

下面给出随机变量函数的期望计算定理:

定理3.1. 设 X 为离散型随机变量, $g: \mathbb{R} \rightarrow \mathbb{R}$ 是连续函数, 随机变量 Y 是 X 的函数: $Y = g(X)$. 若 X 的分布列为 $P(X = x_k) = p_k$ ($k = 1, 2, \dots$), 且 $\sum_k g(x_k)p_k$ 绝对收敛, 则有

$$E[Y] = E[g(X)] = \sum_{k=1}^{\infty} g(x_k)p_k.$$

此定理意义在于计算期望 $E(Y)$ 时, 不必计算 Y 的分布律, 只须利用 X 的分布律即可.

Proof. 证明的本质是利用绝对收敛保证无穷级数任意重排的级数仍然收敛于原无穷级数的和. 首先给出 X 和 Y 的分布列.

X	x_1	x_2	\cdots	x_n	\cdots
P	p_1	p_2	\cdots	p_n	\cdots
Y	y_1	y_2	\cdots	y_n	\cdots
P	$P(Y = y_1)$	$P(Y = y_2)$	\cdots	$P(Y = y_n)$	\cdots

将 $x_1, x_2, \dots, x_n, \dots$ 进行重新分组

$$\underbrace{x_{1,1}, x_{1,2}, \dots, x_{1,k_1}}_{y_1=g(x_{1,j}) \ (j \in [k_1])}, \underbrace{x_{2,1}, x_{2,2}, \dots, x_{2,k_2}}_{y_2=g(x_{2,j}) \ (j \in [k_2])}, \dots, \underbrace{x_{n,1}, x_{n,2}, \dots, x_{n,k_n}}_{y_n=g(x_{n,j}) \ (j \in [k_n])}, \dots$$

由此可得 Y 的分布列为

$$P[Y = y_i] = \sum_{j=1}^{k_i} p_{i,j}$$

从而得到

$$E[Y] = \sum_{i=1}^{\infty} y_i P[Y = y_i] = \sum_{i=1}^{\infty} y_i \sum_{j=1}^{k_i} p_{i,j} = \sum_{i=1}^{\infty} \sum_{j=1}^{k_i} g(x_{i,j}) p_{i,j} = \sum_{j=1}^{\infty} g(x_j) p_j,$$

最后一个等式成立是因为绝对收敛级数的重排. □

推论3.1. 设 X 为离散型随机变量, 其分布列为 $p_k = P(X = x_k)$. $g_i : \mathbb{R} \rightarrow \mathbb{R}$ 是连续函数($i \in [n]$), 且 $E(g_i(X))$ 存在. 对任意常数 c_1, c_2, \dots, c_n , 有

$$E(c_1 g_1(X) + c_2 g_2(X) + \cdots + c_n g_n(X)) = \sum_{i=1}^n c_i E(g_i(X)).$$

Proof. 根据定理 3.1有

$$\begin{aligned} & E(c_1 g_1(X) + c_2 g_2(X) + \cdots + c_n g_n(X)) \\ &= \sum_{k=1}^{\infty} p_k (c_1 g_1(x_k) + c_2 g_2(x_k) + \cdots + c_n g_n(x_k)) \\ &= \sum_{i=1}^n c_i \sum_{k=1}^{\infty} p_k g_i(x_k) = \sum_{i=1}^n c_i E(g_i(X)). \end{aligned}$$

□

给定随机变量 X 和连续函数 $g : \mathbb{R} \rightarrow \mathbb{R}$, $E[g(X)]$ 与 $g(E[X])$ 有完全不同的涵义, 没有直接的大小关系, 但对于特殊的凸函数或凹函数, 有一个重要的结论.

定义3.3. 函数 $g : (a, b) \rightarrow \mathbb{R}$ 满足:

- 对 $\forall x_1, x_2 \in (a, b)$, 有 $g(x_1/2 + x_2/2) \leq (g(x_1) + g(x_2))/2$, 称函数 g 是定义在 (a, b) 上的凸函数;
- 对 $\forall x_1, x_2 \in (a, b)$, 有 $g(x_1/2 + x_2/2) \geq (g(x_1) + g(x_2))/2$, 称函数 g 是定义在 (a, b) 上的凹函数.

定理3.2 (Jensen不等式). 设 X 为随机变量, 若 $g: \mathbb{R} \rightarrow \mathbb{R}$ 是连续的凸函数, 则有 $g(E(X)) \leq E(g(X))$; 若 $g: \mathbb{R} \rightarrow \mathbb{R}$ 是连续的凹函数, 则有 $g(E(X)) \geq E(g(X))$.

此定理对人工智能或计算机学科的分析非常重要, 其证明略去. 根据此定理, 可以直接得到:

- 对任意随机变量 X , 有 $(E(X))^2 \leq E(X^2)$;
- 对任意随机变量 X , 有 $e^{E(X)} \leq E(e^X)$.

3.2.2 方差

数学期望反映了 X 取值的平均值, 而方差则反映了 X 偏离平均值 $E(X)$ 的程度. 例如三个随机变量 X, Y, Z , 其分布列分别为

$$P(X=0)=1; \quad P(Y=1)=1/2, P(Y=-1)=1/2; \quad P(Z=2)=1/5, P(Z=-1/2)=4/5.$$

尽管 $EX = EY = EZ = 0$, 即三随机变量的均值相同, 但这三个随机变量与均值0的偏离程度有很大的差异, 即本节所研究的方差.

定义3.4. 设 X 为离散性随机变量, 其分布列为 $p_k = P(X = x_k) \quad k = 1, 2, \dots$, 若 $E(X) = \sum_k x_k p_k$ 存在, 以及 $E(X - E(X))^2 = \sum_k p_k (x_k - E(X))^2$ 存在, 称 $E(X - E(X))^2$ 为随机变量 X 的方差 (variance), 记为 $Var(X)$ 或 $D(X)$, 即

$$Var(X) = D(X) = E(X - E(X))^2 = \sum_k p_k (x_k - E(X))^2 = \sum_k p_k \left(x_k - \sum_k x_k p_k \right)^2.$$

称 $\sqrt{Var(X)}$ 为标准差 (standard deviation), 记为 $\sigma(X)$.

方差有另外一种简化计算的定义:

定理3.3. 对随机变量 X , 有

$$Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2.$$

Proof. 根据定义和推论 3.1 有

$$\begin{aligned} Var(X) &= E(X - E(X))^2 = E(X^2 - 2XE(X) + E^2(X)) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2. \end{aligned}$$

□

例3.7. 随机变量 X 的取值为 x_1, x_2, \dots, x_n , 其分布列为 $P(X = x_i) = 1/n$. 求计算随机变量 X 方差需要遍历 x_1, x_2, \dots, x_n 几遍.

解. 若利用定义 $\text{Var}(X) = E(X - E(X))^2$ 需要遍历 x_1, x_2, \dots, x_n 两遍, 若利用 $\text{Var}(X) = E(X^2) - (E(X))^2$ 则只需要遍历 x_1, x_2, \dots, x_n 一遍. \square

下面看看方差的性质:

性质3.3. i) 如果随机变量 $X \equiv c$, 则有 $\text{Var}(X) = 0$;

ii) 对任意随机变量 X 和常数 a, b , 有 $\text{Var}(aX) = a^2 \text{Var}(X)$ 以及 $\text{Var}(aX + b) = a^2 \text{Var}(X)$;

iii) 对任意随机变量 X 和常数 c , 有 $\text{Var}(X) = E(X - E(X))^2 \leq E(X - c)^2$.

Proof. 如果随机变量 $X \equiv c$, 则有 $E(X) = c$, 从而得到 $\text{Var}(X) = E(X - E(X))^2 = E(c - c)^2 = 0$.

对于ii), 有 $E(aX) = aE(X)$ 和 $E(aX + b) = aE(X) + b$, 以及

$$\begin{aligned}\text{Var}(aX) &= E(aX - E(aX))^2 = a^2 E(X - E(X))^2 = a^2 \text{Var}(X), \\ \text{Var}(aX + b) &= E(aX + b - E(aX + b))^2 = a^2 E(X - E(X))^2 = a^2 \text{Var}(X).\end{aligned}$$

对于iii), 有

$$\begin{aligned}E(X - c)^2 &= E(X - E(X) + E(X) - c)^2 \\ &= E(X - E(X))^2 + E[(X - E(X))(E(X) - c)] + (E(X) - c)^2 \\ &= E(X - E(X))^2 + (E(X) - c)^2 \geq E(X - E(X))^2.\end{aligned}$$

\square