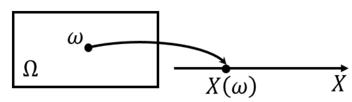
第 3 章　离散型随机变量

有些随机试验的结果是数值的, 例如, 抛一枚骰子的点数分别为 $1, 2, \cdots, 6$; 国家一年出生的婴儿数分别为 $0, 1, 2, \cdots, n, \cdots$. 有些试验结果可能与数值, 但可以用数值来表示, 例如, 抛一枚硬币, 用 0 表示 '正面朝上', 用 1 表示 '正面朝下'. 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时, 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 与一个实数 $X(\omega)$ 相对应, $X(\omega)$ 是 ω 的实值函数, 称实值函数 $X(\omega): \Omega \to \mathbb{R}$ 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, 用随机变量 X 表示出现的点数, 则随机变量 $X \in [6]$. 出现的点数不超过 4 的事件可表示为 $\{X \leqslant 4\}$; 出现偶数点的事件可表示为 $\{X = 2, 4, 6\}$.
- 用随机变量 X 表示一盏电灯的寿命, 其取值为 $[0, +\infty)$, 电灯寿命不超过 500 小时的事件可表示为 $\{X \leqslant 500\}$.

通过随机变量来形式化描述随机现象或随机事件, 从而利用数学工具来研究概率, 例如 $\{X \leqslant -\infty\}$ 表示不可能事件, 以及 $\{X \leqslant +\infty\}$ 表示必然事件.

根据随机变量的取值, 可分为离散型随机变量和连续型随机变量. 若随机变量 X 的取值是有限的、或无限可列的, 则称 X 为 **离散型随机变量**; 若随机变量 X 的取值是无限不可列的, 则称 X 为 **非离散型随机变量**. 本章主要研究离散型随机变量.

3.1 离散型随机变量及分布列

离散型随机变量 X 的取值是有限或无限可列的, 不妨假设其取值为 $x_1, x_2, \cdots, x_n, \cdots$, 事件 $\{X = x_k\}$ 的概率记为

$$p_k = P(X = x_k), \quad k = 1, 2, \cdots,$$

称之为随机变量 X 的 **分布列**.

分布列包含了随机变量的取值和概率, 从而完整地刻画了离散随机变量的概率属性, 也可以用表格表示分布列, 如下所示:

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

根据概率的非负性和完备性有

性质 3.1 随机变量 X 的分布列 $p_k = P(X = x_k)$ $(k \geqslant 1)$ 满足 $p_k \geqslant 0$ 和 $\sum_k p_k = 1$.

下面来看看一些离散随机变量的例子:

例 3.1 设随机变量 X 的分布列 $P(X = k) = c/4^k$ $(k = 0, 1, 2, \cdots)$, 求概率 $P(X = 1)$.

解 根据概率的完备性有

$$1 = \sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{c}{4^k} = \frac{4}{3}c,$$

求解得到 $c = 3/4$, 进一步有 $P(X = 1) = 3/16$.

例 3.2 给定常数 $\lambda > 0$, 随机变量 X 的分布列 $p_i = c\lambda^i/i!$ $(i \geqslant 0)$, 求 $P(X > 2)$.

解 根据概率的完备性有

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

从而得到 $c = e^{-\lambda}$, 进一步得到

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

例 3.3 从 $\{1, 2, \ldots, 10\}$ 中不放回随机任意取 5 个数, 令随机变量 X 表示所取 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

$$P(X = k) = \binom{k-1}{4} \Big/ \binom{10}{5} \quad (k = 5, 6, \ldots, 10).$$

由此可得 X 的分布列表格为

X	5	6	7	8	9	10
P	1/252	5/252	15/252	35/252	70/252	126/252

3.2 离散型随机变量的期望和方差

随机变量的取值具有一定的随机性, 我们希望研究随机变量的一些不变量, 用以刻画随机变量的特征, 最常见的特征是期望与方差.

3.2.1 期望

定义 3.1 设离散型随机变量 X 的分布列为 $P(X = x_k) = p_k$ $(k \geqslant 1)$, 若级数 $\sum_{k=1}^{\infty} p_k x_k$ 绝对收敛, 称级数和为随机变量 X 的 **期望** (expectation), 又被称为 **均值** (mean) 或 **加权平均** (weighted average), 记为 $E(X)$, 即

$$E(X)=\sum_{k=1}^{\infty}p_kx_k.$$

期望 $E(X)$ 反映随机变量 X 的平均值, 由随机变量的分布列决定, 是常量而不是变量, 其本质是随机变量的取值 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, 因此期望 $E(X)$ 反映了 X 可能值的平均值, 不会随次序改变而改变. 根据随机变量随机变量 X 的分布列可直接计算其期望.

例 3.4　随意掷一枚骰子, X 表示观察到的点数, 求 $E[X]$.

解　随机变量 X 的取值为 $1,2,\cdots,6$, 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 的期望为

$$E(X)=(1+2+3+4+5+6)/6=3.5.$$

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, 同一盒子内的球无顺序关系, 用 X 表示有球盒子的最小号码, 求 $E(X)$.

解　先给出 X 的分布列

$$P(X=1)=\frac{\binom{3}{1}3^2+\binom{3}{2}3+1}{4^3}=\frac{37}{64},\quad P(X=2)=\frac{\binom{3}{1}2^2+\binom{3}{2}2+1}{4^3}=\frac{19}{64},$$
$$P(X=3)=\frac{\binom{3}{1}+\binom{3}{2}+1}{4^3}=\frac{7}{64},\qquad P(X=4)=\frac{1}{64}.$$

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　有 n 把钥匙只有一把能打开门, 随机选取一把试开门, 若打不开则除去, 求打开门需要尝试次数的期望.

解　设随机变量 X 表示尝试开门的次数, 其分布列为

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

进一步可得打开门次数的平均数

$$E(X)=\sum_{k=1}^{n}\frac{k}{n}=\frac{(1+n)n}{2n}=\frac{n+1}{2}.$$

根据期望的定义有如下性质:

性质 3.2 若随机变量 $X \equiv c \in \mathbb{R}$, 则 $E(c) = c$.

性质 3.3 对随机变量 X 和常数 $a, b \in \mathbb{R}$, 有 $E(aX + b) = aE(X) + b$.

证明 设随机变量 X 的分布列为 $P(X = x_k) = p_k$, 则随机变量 $Y = aX + b$ 的分布列为 $P(Y = ax_k + b) = p_k$, 进而有

$$E[aX + b] = \sum_{k\geqslant 1}(ax_k + b)p_k = a\sum_{k\geqslant 1} x_k p_k + b\sum_{k\geqslant 1} p_k = aE[X] + b.$$

对随机变量函数的期望, 有如下定理:

定理 3.1 设离散型随机变量 X 的分布列为 $P(X = x_k) = p_k$ $(k \geqslant 1)$, 若 $g : \mathbb{R} \to \mathbb{R}$ 是连续的函数, 且级数 $\sum_{k\geqslant 1} g(x_k)p_k$ 绝对收敛, 则有

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

根据该定理可知当计算随机变量 $Y = g(X)$ 的期望时, 不需计算 Y 的分布列, 只需利用 X 的分布列即可计算期望 $E[Y]$.

证明 证明的思想是利用绝对收敛保证无穷级数任意重排后的级数仍收敛于原无穷级数的和. 根据题意有 X 的分布列为 $P(X = x_k) = p_k$ 以及随机变量函数 $Y = g(X)$ 有

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	$\cdots$

注意 y_i 可能等于 y_j $(i \neq j)$, 因此 $P(Y = y_j) = p_j$ 不是随机变量 Y 的分布列. 为构造 Y 的分布列, 我们将 $x_1, x_2, \ldots, x_n, \ldots$ 进行重新分组,

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

其中 $y_i' \neq y_j'$ $(i \neq j)$. 由此可得随机变量 Y 的分布列为

$$P[Y = y_i'] = \sum_{j=1}^{k_i} p_{i,j} = \sum_{k\geqslant 1, y_i'=g(x_k)} p_k,$$

进一步得到随机变量 Y 的期望为

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

最后一个等式成立是因为绝对收敛级数重排后其和不变.

推论 3.1 对离散型随机变量 X 和连续函数 $g_i:\mathbb{R}\to\mathbb{R}$ $(i\in[n])$, 若每个函数的期望 $E(g_i(X))$ 存在, 则对任意常数 $c_1,c_2,\ldots,c_n$ 有

$$E\left(c_1g_1(X)+c_2g_2(X)+\cdots+c_ng_n(X)\right)=\sum_{i=1}^{n}c_iE(g_n(X)).$$

证明 根据定理 3.1有

$$\begin{aligned}&E\left(c_1g_1(X)+c_2g_2(X)+\cdots+c_ng_n(X)\right)\\ =&\sum_{k=1}^{\infty}p_k(c_1g_1(x_k)+c_2g_2(x_k)+\cdots+c_ng_n(x_k))\\ =&\sum_{i=1}^{n}c_i\sum_{k=1}^{\infty}p_kg_i(x_k)=\sum_{i=1}^{n}c_iE(g_i(X)).\end{aligned}$$

由此容易得到 $E(X^2+X+\sin X+4)=E(X^2)+E(X)+E(\sin X)+4$.

在实际应用中往往不知道随机变量的分布, 但需要对期望进行一定的估计, 为此需要引入一些不等式. 给定随机变量函数 $Y=g(X)$, 下面探讨 $E(g(X))$ 和 $g(E(X))$ 之间的大小关系.

定义 3.2 若函数 $g\colon[a,b]\to\mathbb{R}$ 对任意 $x_1,x_2\in[a,b]$ 和 $\lambda\in[0,1]$, 有 $g(\lambda x_1+(1-\lambda)x_2)\leqslant\lambda g(x_1)+(1-\lambda)g(x_2)$ 成立, 称函数 $g(x)$ 是定义在 $[a,b]$ 上的 **凸函数**;

若对任意 $x_1,x_2\in[a,b]$ 和 $\lambda\in[0,1]$, 有 $g(\lambda x_1+(1-\lambda)x_2)\geqslant\lambda g(x_1)+(1-\lambda)g(x_2)$ 成立, 则称函数 $g(x)$ 是定义在 $[a,b]$ 上的 **凹函数**.

下面介绍著名的 **琴生不等式** (Jensen's inequality), 常用于各种推导估计.

定理 3.2 对离散型随机变量 $X\in[a,b]$ 和连续凸函数 $g\colon[a,b]\to\mathbb{R}$, 有

$$g(E(X))\leqslant E(g(X));$$

对离散型随机变量 $X\in[a,b]$ 和连续凹函数 $g\colon[a,b]\to\mathbb{R}$, 有

$$g(E(X))\geqslant E(g(X)).$$

证明 为了证明的简洁起见, 这里考虑有限的样本空间和凸函数情况, 可类似考虑其它情况. 设随机变量 X 的取值为 $x_1,x_2,\cdots,x_n$, 其分布列为 $P(X=x_k)=p_k\geqslant0$, 根据概率性质有 $\sum_k p_k=1$. 我们需要证明 $g(E(X))\leqslant E[g(X)]$, 即不等式

$$g(p_1x_1+p_2x_2+\cdots+p_nx_n)\leqslant p_1g(x_1)+p_2g(x_2)+\cdots+p_ng(x_n).\tag{3.1}$$

这里对 n 采用归纳法证明, 当 $n=2$ 时由凸函数的定义直接可证. 不妨假设 $n=m-1$ 时成立 $(m\geqslant 3)$, 下面证明当 $n=m$ 亦成立. 首先有

$$\begin{aligned}
g(p_1x_1+p_2x_2+\cdots+p_mx_m) &= g\left(p_1x_1+(1-p_1)\left[\frac{p_2}{1-p_1}x_2+\cdots+\frac{p_m}{1-p_1}x_m\right]\right) \\
&\leqslant p_1g(x_1)+(1-p_1)g\left(\frac{p_2}{1-p_1}x_2+\cdots+\frac{p_m}{1-p_1}x_m\right)
\end{aligned}$$

这里利用 $g(p_1x_1+(1-p_1)x_1')\leqslant p_1g(x_1)+(1-p_1)g(x_1')$, 其中 $x_1'=(x_2p_2+\cdots+x_mp_m)/(1-p_1)$. 容易发现 $p_i/(1-p_1)\geqslant 0$ 且 $\sum_{i=2}^m p_i/(1-p_1)=1$, 根据归纳假设有

$$g\left(\frac{p_2}{1-p_1}x_2+\cdots+\frac{p_m}{1-p_1}x_m\right)\leqslant\frac{p_2}{1-p_1}g(x_2)+\cdots+\frac{p_m}{1-p_1}g(x_m),$$

代入即可完成证明.

对任意离散型随机变量 X, 根据 Jensen 不等式有

$$(E(X))^2\leqslant E(X^2)\qquad \text{和} \qquad e^{E(X)}\leqslant E(e^X).$$

3.2.2 方差

数学期望反映了 X 取值的平均值, 对三个随机变量 X, Y 和 Z, 其分布列分别为

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

尽管随机变量的均值相同 $EX=EY=EZ=0$, 但这三个随机变量与期望的偏离程度有很大的差异, 本节研究随机变量 X 与期望 $E(X)$ 的偏离程度, 即方差.

定义 3.3 离散性随机变量 X 的分布列为 $p_k=P(X=x_k)$ $(k\geqslant 0)$, 若期望 $E(X)=\sum_k x_kp_k$ 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 $\mathrm{Var}(X)$ 或 $D(X)$, 即

$$\mathrm{Var}(X)=\mathrm{D}(X)=E(X-E(X))^2=\sum_k p_k(x_k-E(X))^2=\sum_k p_k\left(x_k-\sum_k x_kp_k\right)^2,$$

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

根据方差的绝对收敛性可知方差不会随随机变量取值的顺序改变而改变, 进而根据期望的性质有

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