

9.2

二项分布

9.2.1 二项分布的数学表示

二项分布是生活中会遇到的另外一种随机变量分布。例如我们去参加球赛，一共参加了 10 场比赛，赢的场次会是多少呢？这个结果不好说。但如果我们知道每场赢的概率，就大致能够估计赢的场次。例如每场赢的概率为 0.5，直觉上会有这样一个估计：赢 5 场左右的概率比较大，赢 9 场或只赢 1 场的概率会比较小。如果是新手，每场赢的概率可能比较小，例如为 0.2，这时要估计 10 场比赛赢的场次，直觉上会有这样一个估计：很可能赢的场次比较少，不太可能赢很多场次。对上面的例子加上一些限定条件，就得到了二项分布：假定每次比赛结果只有赢和输两种情况；每次比赛的结果是独立的，即当前场次的输赢不受前一次比赛结果的影响；一共参加了 n 场比赛。此时赢的次数就是一个符合二项分布的随机变量。

在概率论和统计学中，二项分布是 n 次独立的“成功 / 失败”试验中成功次数的离散概率分布，即 n 重伯努利试验的概率分布。二项分布有如下属性。

- (1) 二项分布中的每个试验都是独立的。
- (2) 在试验中只有两个可能的结果：成功或失败。
- (3) 总共进行了 n 次试验。
- (4) 所有试验成功或失败的概率是相同的，即试验是一样的。

根据以上属性，给出二项分布的数学定义。

定义 9.4 若随机变量 X 的概率分布为

$$P\{X=k\}=C_n^k p^k q^{n-k}, k=0,1,2,\dots,n \quad (9-1)$$

其中 $0 < p < 1, q = 1 - p$ ，则称 X 服从参数为 n, p 的二项分布，记作 $X \sim B(n, p)$ 。上式中， p 表示单次试验成功的概率， q 表示单次试验失败的概率，满足 $p + q = 1$ ； n 表示试验次数，当 $n=1$ 时，二项分布就还原为 (0-1) 分布。

公式 (9-1) 表示了离散随机变量 X 在取 k 值时的概率，这种表示离散随机变量在各个特定值的概率的函数称为概率质量函数 (Probability Mass Function, PMF)。对于二项分布，概率质量函数的均值实际上就是 $n \cdot p$ 。

对应连续型随机变量，二项分布也有累积概率密度函数 (Cumulative Distribution Function, CDF) 的概念。如公式 (9-1) 表示随机变量 X 服从二项分布，此时有如下公式，则这个公式可以作为二项分布的累计概率密度函数的计算公式。