选做题: 二项分布与超几何分布的期望

2023 年 6 月 2 日

我们在课堂上补充了期望的**线性**性质:

0.1.定理. 设 X,Y 为 (同一个概率空间上的) 随机变量, 且期望 $\mathbb{E}[X]$,$\mathbb{E}[Y]$ 均存在, 则:

1. 随机变量 $X+Y$ 的期望存在, 并且

$$\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y].$$

2. 对任意 $\lambda\in\mathbb{R}$, 随机变量 λX 的期望存在, 并且

$$\mathbb{E}[\lambda X]=\lambda\mathbb{E}[X].$$

虽然在课堂上只对样本空间 Ω 为有限集的特殊情形做出简要说明, 但实际上此定理对任何随机变量 (例如连续型随机变量) 也都成立. 此定理 (尤其是第 1 条) 表明: 为计算某个随机变量 X 的期望, 可以把 X 拆成若干个更简单的随机变量之和, 对每个简单的随机变量分别求期望再相加即可. 这次选做作业将用此方法求出二项分布、超几何分布的期望, 并给出更多类似问题作为练习.

0.2.习题.(二项分布的期望) 已知随机变量 X 服从参数为 n,p 的二项分布, 即 $X\sim B(n,p)$, 此题将证明: $\mathbb{E}[X]=np$.

1. **方法一**: 用分布列来计算. 二项分布的分布列众所周知: 对任意 $0\leqslant k\leqslant n$, $\mathbb{P}(X=k)=C_n^kp^k(1-p)^{n-k}$, 因此

$$\mathbb{E}[X]=\sum_{k=0}^{n}kC_n^kp^k(1-p)^{n-k}.$$

   下面我们将直接计算上式右边, 从而求得 X 的期望.

   (1) 对任意 $1\leqslant k\leqslant n$, 直接验证 $kC_n^k=nC_{n-1}^{k-1}$.

   (2) 利用上一小问的结果以及二项式定理, 证明:

$$\sum_{k=0}^{n}kC_n^kp^k(1-p)^{n-k}=np,$$

   从而 $\mathbb{E}[X]=np$.

2. **方法二**: 利用期望的线性性质. 回忆二项分布的实际意义: n 次独立重复试验, 每次试验成功的概率都为 p, 则 X 为这 n 次试验当中成功的总次数.

   (1) 对每个 $1 \leqslant k \leqslant n$, 定义随机变量 X_k 如下:

   $$X_k = \begin{cases} 1 & \text{若第 } k \text{ 次试验成功}, \\ 0 & \text{若第 } k \text{ 次试验失败}. \end{cases}$$

   写出随机变量 X_k 的分布列, 并验证 $\mathbb{E}[X_k] = p$, $(1 \leqslant k \leqslant n)$.

   (2) 注意 $X = X_1 + X_2 + \cdots + X_n$, 从而验证 $\mathbb{E}[X] = np$.

0.3. 习题.(超几何分布的期望) 已知随机变量 X 服从参数为 N, n, M 的超几何分布, 即 $X \sim H(N, n, M)$, 本题将证明: $\mathbb{E}[X] = \frac{nM}{N}$.

1. **方法一**: 用分布列来计算. 二项分布的分布列众所周知: 对于 $0 \leqslant k \leqslant n$, $\mathbb{P}(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$, 从而 X 的期望

   $$\mathbb{E}[X] = \sum_{k=0}^{n} k \cdot \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}.$$

   接下来只需要计算上式右边. 如果你们对组合数、组合恒等式非常熟悉, 其实可以强行暴力计算出来, 但这并不是本次作业的重点, 故从略.

2. **方法二**: 利用期望的线性性质. 回忆超几何分布的实际意义: 总共有 N 个球, 包括 M 个红球 (好东西) 与 $(N - M)$ 个黑球 (坏东西), 从这 N 个球中随机抓取 n 个球, 这 n 个球之中的红球个数为 X.

   (1) 记 $R = \{r_1, r_2, ..., r_M\}$ 为这 M 个红球之全体. 对每个 $1 \leqslant k \leqslant M$, 定义随机变量 X_k 如下:

   $$X_k = \begin{cases} 1 & \text{若随机抓取的 } n \text{ 个球中包括红球 } r_k, \\ 0 & \text{其余情况}. \end{cases}$$

   对每个 k, 写出 X_k 的分布列, 并验证 $\mathbb{E}[X_k] = \frac{n}{N}$.

   (2) 验证 $X = X_1 + X_2 + \cdots + X_M$, 从而求得 $\mathbb{E}[X] = \frac{nM}{N}$.

0.4. 习题. 回忆之前做过的一道随堂练习题: 掷一枚标准的骰子, 若掷出 1, 2, 3 点则得 0 分, 掷出 4 或 5 点得 1 分, 掷出 6 点得 2 分; 现在独立地掷骰子 3 次, 这 3 次所得的总分记作 X. 这次将计算随机变量 X 的期望.

1. **方法一**: 利用分布列来计算 $\mathbb{E}[X]$. 我们之前计算过 X 的分布列, 如下:

X	0	1	2	3	4	5	6
$\mathbb{P}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{7}{24}$	$\frac{11}{54}$	$\frac{7}{72}$	$\frac{1}{36}$	$\frac{1}{216}$

利用上述分布列, 直接计算期望 $\mathbb{E}[X]$.

2. **方法二**: 利用期望的线性性质.

(1) 对于 $k = 1, 2, 3$, 记 X_k 为第 k 次掷骰子所获得的分数. 写出 X_k 的分布列, 并计算 $\mathbb{E}[X]$.

(2) 注意 $X = X_1 + X_2 + X_3$, 从而由期望的线性性质计算 $\mathbb{E}[X]$.