

第3节 二项分布与超几何分布 (★★★)

内容提要

二项分布与超几何分布是两个容易混淆的概念,本节归纳与之相关的一些常见题型,下面先梳理二项分布、超几何分布的概念.

1. 二项分布: 在 n 重伯努利试验中, 设每次试验中事件 A 发生的概率为 $p(0 < p < 1)$, 用 X 表示事件 A 发生的次数, 则 X 的分布列为 $P(X=k) = C_n^k p^k (1-p)^{n-k}$, 其中 $k=0,1,2,\dots,n$, 称随机变量 X 服从二项分布, 记作 $X \sim B(n, p)$.

2. 期望和方差: 若 $X \sim B(n, p)$, 则 $E(X) = np$, $D(X) = np(1-p)$.

3. 超几何分布: 设一批产品共有 N 件, 其中次品 M 件, 其余为合格品, 从 N 件产品中随机抽取 n 件, 记取到次品的件数为随机变量 X , 则 $P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$, 其中 $k=m, m+1, m+2, \dots, r$, 且

$m = \max\{0, n - N + M\}$, $r = \min\{n, M\}$. 具有上述概率分布的随机变量 X 即为服从超几何分布的随机变量, 其均值 $E(X) = n \cdot \frac{M}{N} = np$, 其中 $p = \frac{M}{N}$ 表示抽取一件产品取到次品的概率. 上面的表述较为抽象, 可结合例4的几道题来理解.

4. 二项分布与超几何分布的关系: 对于不放回的抽取, 当 n 远小于 N 时, 每抽取一次后, 对 N 的影响很小, 此时, 超几何分布可用二项分布近似.

典型例题

类型 I: 二项分布概念题

【例1】(2017·新课标II卷) 一批产品的二等品率为0.02, 从这批产品中每次随机取一件, 有放回地抽取100次, X 表示抽到的二等品件数, 则 $D(X) = \underline{\hspace{2cm}}$.

解析: 有放回地抽取, 则下一次抽取不受上一次抽取结果的影响, 故抽取100次可看成100次独立重复试验, 成功次数(抽到二等品的件数)应服从二项分布,

由题意, 二等品率为0.02, 即每次试验的成功概率为0.02, 所以 $X \sim B(100, 0.02)$,

故 $D(X) = 100 \times 0.02 \times (1 - 0.02) = 1.96$.

答案: 1.96

【变式1】假设苏州肯帝亚球队在某赛季的任一场比赛中输球的概率都等于 $p(0 < p < 1)$, 且各场比赛互不影响. 令 X 表示连续9场比赛中输球的场数, 且 $P(X=5) = P(X=6)$, 则球队在这连续9场比赛中输球场数的期望为 $\underline{\hspace{2cm}}$.

解析: 9场比赛可看成9次独立试验, 每场比赛输球的概率相同, 满足重复性, 所以这是独立重复试验, 故输球的场数服从二项分布, 可用其概率分布计算 $P(X=5)$ 和 $P(X=6)$, 从而建立方程求 p ,

由题意, 9场比赛中输球的场数 $X \sim B(9, p)$, 所以 $P(X=5) = C_9^5 p^5 (1-p)^4$, $P(X=6) = C_9^6 p^6 (1-p)^3$,

故 $P(X=5)=P(X=6)$ 即为 $C_9^5 p^5 (1-p)^4 = C_9^6 p^6 (1-p)^3$, 解得: $p = \frac{3}{5}$, 所以 $E(X) = 9p = \frac{27}{5}$.

答案: $\frac{27}{5}$

【变式 2】 甲与乙进行投篮游戏, 在每局游戏中两人分别投篮两次, 每局投进的次数之和不小于 3 则该局游戏胜利, 已知甲、乙两名队员投篮相互独立且投进的概率均为 $\frac{2}{3}$, 现进行 27 局游戏, 设 X 为甲、乙两名队员胜利的局数, 则 X 的期望为_____.

解析: 27 局游戏可看成 27 重伯努利试验, 胜利的次数 X 服从二项分布, 故先求一局游戏胜利的概率,

由于甲、乙每球投进的概率都是 $\frac{2}{3}$, 所以各投两球可看成共投 4 球, 由题意, 至少进 3 球即成功,

投的这 4 球又可看成 4 重伯努利试验, 4 次投篮中投进的次数也服从二项分布,

设投进的次数为 Y , 则 $Y \sim B(4, \frac{2}{3})$, 所以 $P(Y \geq 3) = P(Y=3) + P(Y=4) = C_4^3 \times (\frac{2}{3})^3 \times (1 - \frac{2}{3}) + C_4^4 \times (\frac{2}{3})^4 = \frac{16}{27}$,

即一局游戏胜利的概率为 $\frac{16}{27}$, 所以 $X \sim B(27, \frac{16}{27})$, 故 $E(X) = 27 \times \frac{16}{27} = 16$.

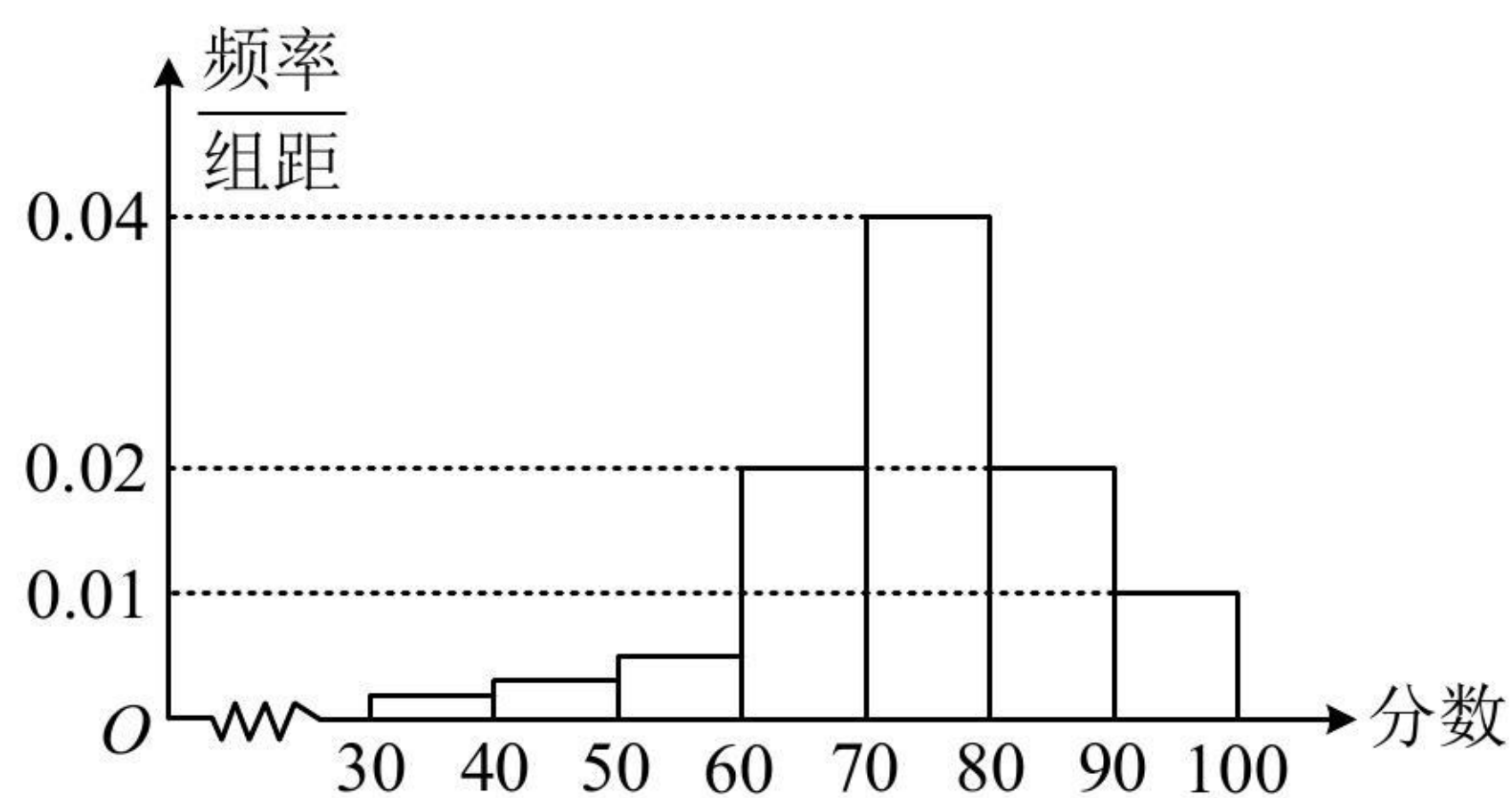
答案: 16

【反思】 在分析 n 次独立重复试验时, 一定要先弄清楚一次试验中的成功概率, 一次试验不一定是投一次篮, 或抛一枚硬币, 也可能是投多次篮, 或抛几次硬币, 要看题干如何规定.

《一数·高考数学核心方法》

类型 II: 二项分布解答题

【例 2】 某学校高三年级学生参加某项体育测试, 根据男女学生人数比例, 使用分层抽样的方法从中抽取了 100 名学生, 记录他们的分数, 将数据分成 7 组: $[30,40)$, $[40,50)$, \dots , $[90,100]$, 整理得到如下的频率分布直方图:



- (1) 若规定小于 60 分为“不及格”, 从该学校高三年级学生中随机抽取 1 人, 估计该学生不及格的概率;
- (2) 若规定分数在 $[80,90)$ 为“良好”, $[90,100]$ 为“优秀”, 用频率估计概率, 从该校高三年级 (总人数较多) 随机抽取 3 人, 记该项测试分数为“良好”或“优秀”的人数为 X , 求 X 的分布列, 期望和方差.

解: (1) 由图可估计抽到不及格学生的概率为 $1 - 10 \times (0.02 + 0.04 + 0.02 + 0.01) = 0.1$.

(2) 由图可知从该学校高三年级学生中随机抽 1 人, 抽到“良好”的概率为 $10 \times 0.02 = 0.2$, 抽到“优秀”的概率为 $10 \times 0.01 = 0.1$, 所以抽到“良好”或“优秀”的概率为 $0.2 + 0.1 = 0.3$,

(由于这3人是从整个高三年级中抽取的,总人数较多,抽取人数远小于总人数,所以每抽取1人后,下次再抽时,抽到“优秀”或“良好”的概率几乎不变,故可按二项分布处理)

由题意, $X \sim B(3, 0.3)$, 所以 $P(X=0) = C_3^0 \times (1-0.3)^3 = 0.343$, $P(X=1) = C_3^1 \times 0.3 \times (1-0.3)^2 = 0.441$,

$P(X=2) = C_3^2 \times 0.3^2 \times (1-0.3) = 0.189$, $P(X=3) = C_3^3 \times 0.3^3 = 0.027$, 故 X 的分布列为:

X	0	1	2	3
P	0.343	0.441	0.189	0.027

因为 $X \sim B(3, 0.3)$, 所以 $E(X) = 3 \times 0.3 = 0.9$, $D(X) = 3 \times 0.3 \times (1-0.3) = 0.63$.

【反思】像这种从某处取几个人,求取到某类个体的个数的分布列这种题,一定要注意是从总体中取,还是从个体数较少的样本中取.若是前者,由于抽取的人数往往远小于总人数,所以常按二项分布处理;而后者,则按超几何分布来求分布列.

【例3】一款击鼓小游戏的规则如下,每盘游戏都需击鼓三次,每次击鼓要么出现一次音乐,要么不出现音乐,每盘游戏击鼓三次后,出现一次音乐获得10分,出现两次音乐获得20分,出现三次音乐获得100分,没有出现音乐则扣除200分,设每次击鼓出现音乐的概率为 $\frac{1}{2}$,且各次击鼓出现音乐与否相互独立.

- (1) 设每盘游戏获得的分数为 X , 求 X 的分布列;
- (2) 玩三盘游戏,至少有一盘出现音乐的概率是多少;
- (3) 玩过这款游戏的人都发现,若干盘游戏后,与最初的分数相比,分数没有增加反而减少了,请用概率统计的知识分析分数减少的原因.

解: (1) (得分 X 由出现音乐的次数 Y 决定,显然 Y 服从二项分布,故将 X 与 Y 对应起来即可)

设每盘游戏出现音乐的次数为 Y , 则 $Y \sim B(3, \frac{1}{2})$,

由题意,当 Y 分别取 0, 1, 2, 3 时, X 的值对应为 -200, 10, 20, 100,

所以 $P(X=-200) = P(Y=0) = C_3^0 \times (1-\frac{1}{2})^3 = \frac{1}{8}$, $P(X=10) = P(Y=1) = C_3^1 \times \frac{1}{2} \times (1-\frac{1}{2})^2 = \frac{3}{8}$,

$P(X=20) = P(Y=2) = C_3^2 \times (\frac{1}{2})^2 \times (1-\frac{1}{2}) = \frac{3}{8}$, $P(X=100) = P(Y=3) = C_3^3 \times (\frac{1}{2})^3 = \frac{1}{8}$,

所以 X 的分布列为

X	-200	10	20	100
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

(2) (玩三盘游戏,可看成三次独立重复试验,出现音乐的盘数也服从二项分布,我们先算成功概率)

由(1)知玩一盘游戏,不出现音乐的概率为 $\frac{1}{8}$,则玩三盘游戏,不出现音乐的盘数 $Z \sim B(3, \frac{1}{8})$,

(至少有一盘出现音乐情况较多,而其对立事件只有三盘都不出现音乐1种,故用对立事件求概率)

三盘都不出现音乐的概率为 $P(Z=3) = C_3^3 \times (\frac{1}{8})^3 = \frac{1}{512}$, 故所求概率为 $1 - (\frac{1}{8})^3 = \frac{511}{512}$.

(3) (多次游戏后,得分会向均值靠拢,故用随机变量 X 的期望来回答这一问题)

由(1)得 $E(X) = (-200) \times \frac{1}{8} + 10 \times \frac{3}{8} + 20 \times \frac{3}{8} + 100 \times \frac{1}{8} = -\frac{5}{4} < 0$, 所以多次游戏后分数减少的可能性更大.

类型III: 超几何分布概念题

【例4】现有10件产品, 其中有2件次品, 其余为合格品, 从中任取2件, 记抽到次品的件数为 X , 则 $E(X) =$ _____.

解析: 从10件产品中取2件, 取到次品的件数服从超几何分布, 可按古典概率算分布列, 需注意此处次品件数与合格品件数都不少于抽取件数,

由题意, X 可能的取值为0, 1, 2, 且 $P(X=0) = \frac{C_8^2}{C_{10}^2} = \frac{28}{45}$, $P(X=1) = \frac{C_8^1 C_2^1}{C_{10}^2} = \frac{16}{45}$, $P(X=2) = \frac{C_2^2}{C_{10}^2} = \frac{1}{45}$,

所以 $E(X) = 0 \times \frac{28}{45} + 1 \times \frac{16}{45} + 2 \times \frac{1}{45} = \frac{2}{5}$.

答案: $\frac{2}{5}$

【反思】若熟悉超几何分布的期望公式 $E(X) = n \cdot \frac{M}{N}$, 也可速求期望, 本题中, $n=2$, $M=2$, $N=10$.

【变式1】现有10件产品, 其中有1件次品, 其余为合格品, 从中任取2件, 记抽到次品的件数为 X , 则 $E(X) =$ _____.

解析: 次品件数小于抽取件数, 所以 X 不能取完0, 1, 2的所有值, 应先分析 X 的可能取值,

X 可能的取值为0, 1, 且 $P(X=0) = \frac{C_9^2}{C_{10}^2} = \frac{4}{5}$, $P(X=1) = \frac{C_9^1 C_1^1}{C_{10}^2} = \frac{1}{5}$, 故 $E(X) = 0 \times \frac{4}{5} + 1 \times \frac{1}{5} = \frac{1}{5}$.

答案: $\frac{1}{5}$

【反思】也可由 $E(X) = n \cdot \frac{M}{N}$ 求期望, 本题 $n=2$, $M=1$, $N=10$.

【变式2】现有10件产品, 其中有1件合格品, 其余为次品, 从中任取2件, 记抽到次品的件数为 X , 则 $E(X) =$ _____.

解析: 合格品件数小于抽取件数, 所以 X 不能取完0, 1, 2的所有值, 应先分析 X 的可能取值,

由题意, X 可能的取值为1, 2, 且 $P(X=1) = \frac{C_9^1 C_1^1}{C_{10}^2} = \frac{1}{5}$, $P(X=2) = \frac{C_9^2}{C_{10}^2} = \frac{4}{5}$, 所以 $E(X) = 1 \times \frac{1}{5} + 2 \times \frac{4}{5} = \frac{9}{5}$.

答案: $\frac{9}{5}$

【反思】也可由 $E(X) = n \cdot \frac{M}{N}$ 求期望, 本题 $n=2$, $M=9$, $N=10$.

【变式3】现有6件产品, 其中有3件次品, 其余为合格品, 从中任取4件, 记抽到次品的件数为 X , 则 $E(X) =$ _____.

解析：次品、合格品件数都小于抽取件数，所以 X 不能取完 0, 1, 2, 3, 4，应先分析 X 的可能取值，

由题意， X 的可能取值为 1, 2, 3，且 $P(X=1) = \frac{C_3^1 C_3^3}{C_6^4} = \frac{1}{5}$ ， $P(X=2) = \frac{C_3^2 C_3^2}{C_6^4} = \frac{3}{5}$ ， $P(X=3) = \frac{C_3^3 C_3^1}{C_6^4} = \frac{1}{5}$ ，

所以 $E(X) = 1 \times \frac{1}{5} + 2 \times \frac{3}{5} + 3 \times \frac{1}{5} = 2$ 。

答案：2

【总结】变式 3 也可用 $E(X) = n \cdot \frac{M}{N}$ 求期望，所以从上面几道题可以看出，无论次品件数、合格品件数与抽取件数的大小关系如何，超几何分布的分布列都按古典概率计算，且几种情况的期望公式都是

$$E(X) = n \cdot \frac{M}{N}.$$

类型IV：超几何分布综合题

【例 5】在某校举办“青春献礼二十大，强国有我新征程”的知识能力测评中，随机抽查了 100 名学生，其中共有 4 名女生和 3 名男生的成绩在 90 分以上，从这 7 名同学中每次随机抽取 1 人在全校做经验分享，每位同学最多分享一次，记第一次抽到女生为事件 A ，第二次抽到男生为事件 B 。

(1) 求 $P(B|A)$ ；

(2) 若把抽取学生的方式更改为：从这 7 名学生中随机抽取 3 人进行经验分享，记被抽取的 3 人中女生的人数为 X ，求 X 的分布列和数学期望。

解：(1) 在 A 发生的条件下，第二次抽取时，共 3 男 3 女 6 人，此时抽到男生的概率 $P(B|A) = \frac{3}{6} = \frac{1}{2}$ 。

(2) (从 7 名学生中抽 3 人，抽到女生的人数应服从超几何分布，按古典概率计算分布列即可)

由题意， X 可能的取值有 0, 1, 2, 3，且 $P(X=0) = \frac{C_3^3}{C_7^3} = \frac{1}{35}$ ， $P(X=1) = \frac{C_3^2 C_4^1}{C_7^3} = \frac{12}{35}$ ，

$P(X=2) = \frac{C_3^1 C_4^2}{C_7^3} = \frac{18}{35}$ ， $P(X=3) = \frac{C_4^3}{C_7^3} = \frac{4}{35}$ ，所以 X 的分布列为：

X	0	1	2	3
P	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$

故 X 的数学期望 $E(X) = 0 \times \frac{1}{35} + 1 \times \frac{12}{35} + 2 \times \frac{18}{35} + 3 \times \frac{4}{35} = \frac{12}{7}$ 。

【反思】服从超几何分布的随机变量，求期望也可直接套用公式 $E(X) = n \cdot \frac{M}{N}$ ，当然，也可用此公式验证上述结果是否正确。

强化训练

1. (★) 投篮测试中, 已知某同学每次投篮投中的概率为 0.6, 且各次投篮是否投中相互独立, 则该同学投 3 次, 恰好投中 2 次的概率为_____.

2. (2014 • 江西卷 • ★) 10 件产品中有 7 件正品, 3 件次品, 从中任取 4 件, 则恰好取到 1 件次品的概率是_____.

3. (★★) 某人参加一次考试, 共 3 道题, 至少答对其中 2 道才能合格, 若他答对每道题的概率均为 0.6, 则他能合格的概率为_____.

《一数·高考数学核心方法》

4. (2022 • 江苏模拟 • ★★★) 一个盒子里装有大小相同的 4 个黑球和 3 个白球, 从中不放回地取出 3 个球, 则白球个数的数学期望是 ()

- (A) $\frac{4}{7}$ (B) $\frac{9}{7}$ (C) $\frac{12}{7}$ (D) $\frac{16}{7}$

5. (2021 • 天津卷 • ★★★) 甲、乙两人在每次猜谜活动中各猜一个谜语, 若一方猜对且另一方猜错, 则猜对的一方获胜, 否则本次平局. 已知每次活动中, 甲、乙猜对的概率分别为 $\frac{5}{6}$ 和 $\frac{1}{5}$, 且每次活动中甲、乙猜对与否互不影响, 各次活动也互不影响, 则 1 次活动中, 甲获胜的概率为_____; 3 次活动中, 甲至少获胜 2 次的概率为_____.

6. (2023 • 重庆模拟 • ★★★) “锦里开芳宴, 兰缸艳早年”, 元宵节是中国非常重要的传统节日, 某班级准

备进行“元宵福气到”抽奖活动，福袋中装有标号分别为1, 2, 3, 4, 5的五个相同小球，从袋中一次性摸出三个小球，若号码之和是3的倍数，则获奖. 若有5名同学参与此次活动，则恰好3人获奖的概率是 ()

- (A) $\frac{72}{625}$ (B) $\frac{108}{625}$ (C) $\frac{144}{625}$ (D) $\frac{216}{625}$

7. (2022·福建福州模拟·★★★★) 为了保障我国民众的身体健康，某产品在进入市场前必须进行两轮检测，只有两轮检测都合格才能进行销售，否则不能销售. 已知该产品第一轮检测不合格的概率为 $\frac{1}{9}$ ，第二轮检测不合格的概率为 $\frac{1}{10}$ ，两轮检测是否合格相互之间没有影响，若产品可以销售，则每件产品获利40元，若产品不能销售，则每件亏损80元，已知一箱中有尚未检测的4件产品，记该箱产品总共获利 X 元，则 $P(X \geq -80) =$ ()

- (A) $\frac{96}{625}$ (B) $\frac{256}{625}$ (C) $\frac{608}{625}$ (D) $\frac{209}{625}$

《一数·高考数学核心方法》

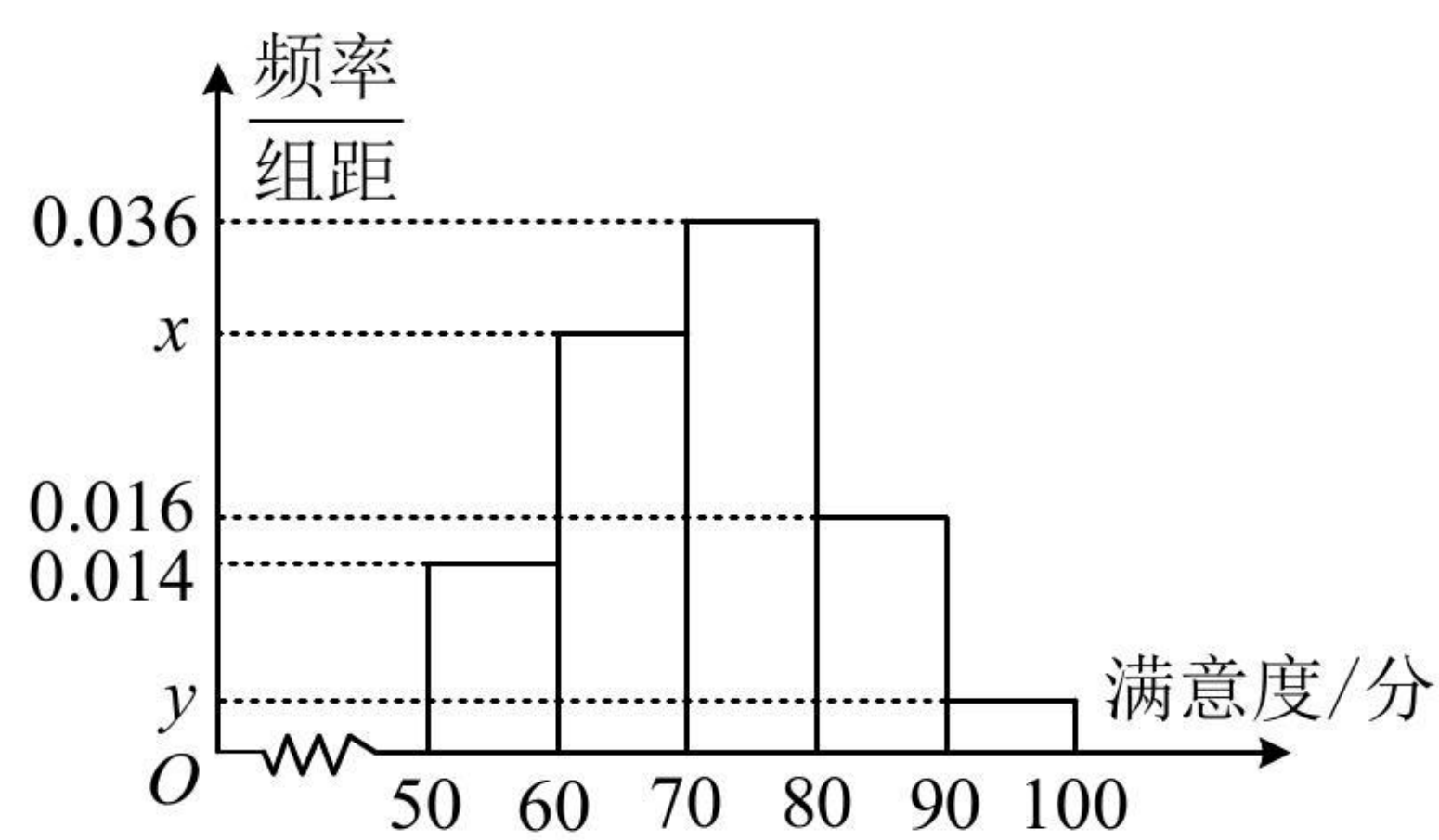
8. (2019·天津卷·★★★★) 设甲、乙两位同学上学期间，每天7:30之前到校的概率均为 $\frac{2}{3}$. 假定甲、乙两位同学到校情况互不影响，且任一同学每天到校情况相互独立.

- (1) 用 X 表示甲同学上学期间的三天中7:30之前到校的天数，求随机变量 X 的分布列和数学期望；
(2) 设 M 为事件“上学期间的三天中，甲同学在7:30之前到校的天数比乙同学在7:30之前到校的天数恰好多2”，求事件 M 发生的概率.

9. (2023·四川成都七中模拟·★★★★) 随着人民生活水平的不断提高，“衣食住行”愈发被人们所重视，

其中对饮食的要求也越来越高. 某地区为了解当地餐饮情况, 随机抽取了 100 人对该地区的餐饮情况进行了问卷调查. 请根据下面尚未完成并有局部污损的频率分布表和频率分布直方图解决下列问题.

组别	分组	频数	频率
第1组	[50,60)	14	0.14
第2组	[60,70)	m	
第3组	[70,80)	36	0.36
第4组	[80,90)		0.16
第5组	[90,100]	4	n



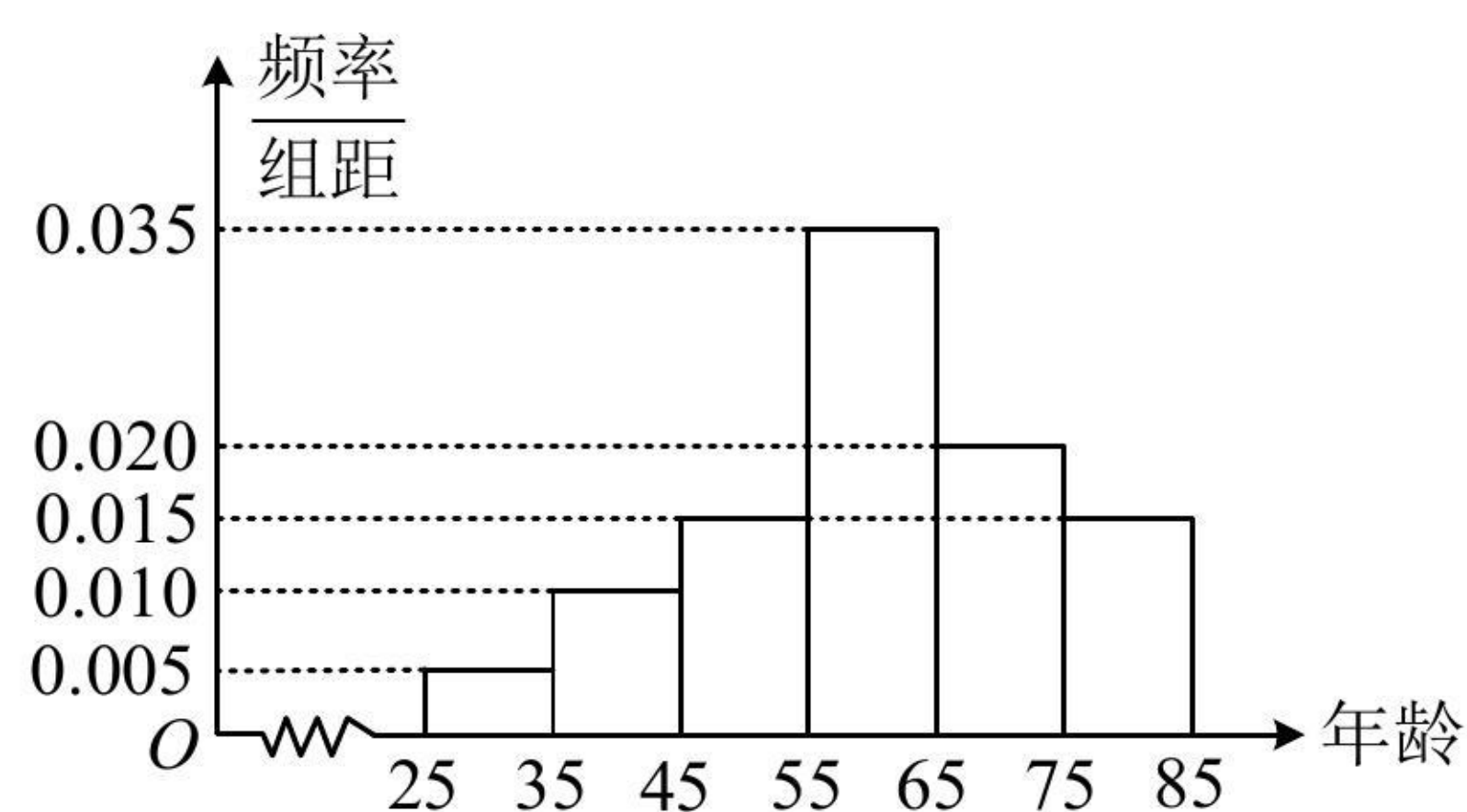
- (1) 求 m, n, x, y 的值;
- (2) 若将满意度在 80 分以上的人群称为“美食客”, 将频率视为概率, 用样本估计总体, 从该地区中随机抽取 3 人, 记其中“美食客”的人数为 ξ , 求 ξ 的分布列和期望.

10. (2018 · 天津卷 (改) · ★★★) 已知某单位甲、乙、丙三个部门的员工人数分别为 24, 16, 16. 现采用按比例分配的分层抽样方法从中抽取 7 人, 进行睡眠时间的调查.

- (1) 应从甲、乙、丙三个部门的员工中分别抽取多少人?
- (2) 若抽出的 7 人中有 4 人睡眠不足, 3 人睡眠充足, 现从这 7 人中随机抽取 3 人做进一步的体检.
 - ①用 X 表示抽取的 3 人中睡眠不足的员工人数, 求随机变量 X 的分布列与数学期望;
 - ②设 A 为事件“抽取的 3 人中, 既有睡眠充足的员工, 也有睡眠不足的员工”, 求事件 A 发生的概率.

11. (2023 · 青海模拟 · ★★★) 2021 年国庆期间, 某县书画协会在县宣传部门的领导下组织了国庆书画

展，参展的 200 幅书画作品反映了该县人民在党的领导下进行国家建设中的艰苦卓绝，这些书画作品的作者的年龄都在 $[25,85]$ 内，根据统计结果，得到如图所示的频率分布直方图：



- (1) 求这 200 位作者年龄的平均数 \bar{x} 和方差 s^2 ；(同一组数据用该区间的中点值作代表)
- (2) 县委宣传部从年龄在 $[35,45)$ 和 $[65,75)$ 的作者中，用按比例分配的分层抽样方法抽取 6 人参加县委组织的表彰大会，现要从 6 人中选出 3 人作为代表发言，设这 3 位发言者的年龄落在区间 $[35,45)$ 的人数为 X ，求 X 的分布列和数学期望.

《一数·高考数学核心方法》

12. (2023·江西模拟·★★★) 党的二十大是全党全国各族人民迈上全面建设社会主义现代化国家的新征程、向第二个百年奋斗目标进军的关键时刻召开的一次十分重要的大会，认真学习宣传和全面贯彻落实党的二十大精神，是当前和今后一个时期的首要政治任务和头等大事. 某校计划举行党的二十大知识竞赛，对前来报名者进行初试，初试合格者进入正赛. 初试有备选题 6 道，从备选题中随机挑出 4 道题进行测试，至少答对 3 道题者视为合格. 已知甲、乙两人报名参加初试，在这 6 道题中甲能答对 4 道，乙能答对每道题的概率均为 $\frac{2}{3}$ ，且甲、乙两人各题是否答对相互独立.

- (1) 分别求甲、乙两人进入正赛的概率；
- (2) 记甲、乙两人中进入正赛的人数为 ξ ，求 ξ 的分布列及 $E(2\xi - 1)$.