

基于思维能力提升的 “二项分布与正态分布”设计示例*

孟 媛(山东省青岛第十六中学)

文章编号:1002-2171(2022)4-0061-04

当前,高三数学第二轮复习中存在着重知识记忆、轻知识逻辑关联,重题型训练、轻思想方法提炼,重解题过程、轻研究对象探究等问题,学生陷入题海训练,学习负担非常大;而且近几年的高考不断创新试题的形式和内涵,聚焦对学生思维水平和关键能力的考查。所以,教师在复习中要创设合理的问题情境,以提升学生数学思维能力为目标,设计逻辑连贯的问题串,注重一题多解、变式教学,使学生在应用知识解决具体问题的过程中理解概念本质、感悟数学思想、学会思考问题。本文结合“二项分布与正态分布”的高三专题复习设计,就课堂上提升学生数学思维、培养学生关键能力的路径进行探讨。

1 教学内容解析

(1)通过具体实例,了解伯努利试验,掌握二项分布及其数字特征,并能解决简单的实际问题;通过误差模型,了解服从正态分布的随机变量;通过具体实例,借助频率分布直方图的几何直观,了解正态分布的特征,了解正态分布的均值、方差及其含义。

(2)常利用情境化试题对二项分布与正态分布知识内容进行考查;题目通常结合古典概型,考查条件概率、独立事件的概率计算,考查学生的数学运算、逻辑推理核心素养;结合 n 次独立重复试验的概念,考查随机变量的二项分布,训练学生的数学抽象、数学建模能力;结合频率分布直方图,考查正态分布曲线的特点、 3σ 原则的应用,训练学生的数据分析、直观想象能力。

2 学习目标

(1)会求二项分布的期望、方差等,了解正态分布曲线的特点、曲线所表示的意义、 3σ 原则等,会求随

机变量在特殊区间内的概率。

(2)能从实际问题中抽象出模型特征,识别二项分布和正态分布,能应用二项分布、正态分布模型解决一些简单的实际问题,发展数学建模素养。

(3)通过主动探究、自主合作、相互交流,从具体事例中抽象出数学模型,充分体会知识的发现过程,及由特殊到一般、由具体到抽象的数学思想方法;体会数学的理性与严谨,了解数学来源于实际、应用于实际的唯物主义思想,形成勇于探索和敢于创新的精神。

3 教学过程

3.1 审清条件,建立二项分布模型

判断一个随机变量是否服从二项分布,关键有两点:一是对立性,即一次试验中,事件发生与否二者必有一;二是重复性,即试验独立重复地进行了 n 次。在根据独立重复试验求二项分布的有关问题时,关键是厘清事件与事件之间的关系,确定二项分布的试验次数 n 和事件发生一次的概率 p ,利用公式 $P(X=k)=C_n^k p^k (1-p)^{n-k} (k=0,1,2,\dots,n)$ 求得概率。

例1 甲、乙两人各有一个装有4个黑球和6个白球的袋子,他们分别从各自的袋子中任取2个球,试求:

(I)甲不放回地抽取,取到黑球数 X 的分布列。

(II)乙有放回地抽取,取到黑球数 Y 的分布列。

(III)记 ξ 为甲、乙两人取到黑球的总数,求随机变量 ξ 的数学期望。

师生活动:学生指出由于甲在取球时是不放回的,因此可以将取球看作是无序的,也就是直接从10个球中取2个球,求其中黑球的个数。学生可以判断该问题服从超几何分布。

(I)由题意知,随机变量 X 服从超几何分布,即 $X \sim H(2,4,10)$,所以随机变量 X 的分布列如表1:

* 本文系山东省教育科学“十三·五”规划2020年度课题“信息技术支撑下的高中数学建模教学实践研究”(课题批准号2020ZC044)研究成果。

表1 随机变量 X 的分布列

X	0	1	2
P	$\frac{1}{3}$	$\frac{8}{15}$	$\frac{2}{15}$

学生指出由于乙在取球时是有放回的,因此每次取球时都是从这10个球中取,即每次取球的情况相同,学生可以判断该问题服从二项分布。

(II)随机变量 Y 服从二项分布,即 $Y \sim B(2, 0.4)$,所以随机变量 Y 的分布列如表2:

表2 随机变量 Y 的分布列

Y	0	1	2
P	$\frac{9}{25}$	$\frac{12}{25}$	$\frac{4}{25}$

第(III)问要求两人一共取到的黑球数 ξ ,应该怎样确定随机变量 ξ 的取值?学生提出,依次列出 ξ 可能的取值不如棋盘法简洁直观,因此列出棋盘如表3:

表3 棋盘法列表

$Y \backslash X$	0	1	2
0	0	1	2
1	1	2	3
2	2	3	4

由棋盘法可得随机变量 ξ 的取值为0,1,2,3,4。

师生活动:我们可否将以上棋盘做进一步改进,使它能表示更多的信息?学生指出可以将概率填入表格,计算每个表格中随机变量 ξ 取值的概率。教师提问:此时概率应该怎样计算?学生认为由于题干表明甲、乙两人分别独立取球,因此互不影响,积事件的概率等于事件概率的积,因此有表4:

表4 改进后的棋盘

$Y \backslash X$	0	1	2
0	$P(X=0, Y=0)$	$P(X=1, Y=0)$	$P(X=2, Y=0)$
1	$P(X=0, Y=1)$	$P(X=1, Y=1)$	$P(X=2, Y=1)$
2	$P(X=0, Y=2)$	$P(X=1, Y=2)$	$P(X=2, Y=2)$

其中由于甲、乙两人互不影响,即随机变量 X 和 Y 相互独立,故 $P(X=i, Y=j) = P(X=i)P(Y=j)$ ($i, j=0, 1, 2$)。

教师点拨:实际上这样列出来的“棋盘”就是二维随机变量 (X, Y) 的联合分布列,利用该联合分布列可以清晰地解决随机变量 $\xi = X + Y$ 的分布情况。

由(I)(II)所得的分布列,列出随机变量 X 和 Y 的联合分布列如表5:

表5 随机变量 X 和 Y 的联合分布列

$Y \backslash X$	0	1	2
0	$\frac{3}{25}$	$\frac{24}{125}$	$\frac{6}{125}$
1	$\frac{4}{25}$	$\frac{32}{125}$	$\frac{8}{125}$
2	$\frac{4}{75}$	$\frac{32}{375}$	$\frac{8}{375}$

整理可得表6:

表6 随机变量 ξ 的分布情况

ξ	0	1	2	3	4
P	$\frac{3}{25}$	$\frac{44}{125}$	$\frac{134}{375}$	$\frac{56}{375}$	$\frac{8}{375}$

$$\text{所以 } E(\xi) = 0 \times \frac{3}{25} + 1 \times \frac{44}{125} + 2 \times \frac{134}{375} + 3 \times \frac{56}{375} + 4 \times \frac{8}{375} = 1.6.$$

设计意图:第(I)(II)问中体现了二项分布和超几何分布在命题上的一个明显特征,即“有放回”和“无放回”,其中“有放回”为二项分布,“无放回”为超几何分布。

当涉及两个离散型随机变量时,可通过棋盘法快速找出随机变量的所有可能取值,将棋盘法进一步优化,实际上就可以得到二维随机变量的联合分布列。由以上分析可以看出,第(III)问中先求分布列再求随机变量的数学期望,运算量较大,过程烦琐。

追问:本题中随机变量之间存在关系 $\xi = X + Y$,据此如何求 ξ 的数学期望?

由数学期望的运算性质可得 $E(\xi = X + Y) = E(X) + E(Y) = np + n \cdot \frac{M}{N} = 1.6$ 。

设计意图:在解答本题的过程中,可以借助两个随机变量的联合分布列使运算过程清晰明了,但是运算过程依然烦琐,故而借助数学期望的运算性质、二项分布以及超几何分布的期望公式,可以快捷地进行计算。

变式1:已知某运动员投篮命中率为 p ($0 < p < 1$),假设他投篮10次恰好命中8次的概率为 $f(p)$ 。

(I)求 $f(p)$ 取得最大值时 p 的取值。

(II)以(I)中确定的 p 值,假设他共投篮25次,恰好命中 k 次的概率为 p_k ,当 k 取何值时 p_k 最大?

师生活动:第(I)问由题意知 $f(p) = C_{10}^8 p^8 (1-p)^2$,求 $f(p)$ 取得最大值时 p 的取值,即将 p 当作自变量研究函数的最值问题,因此可求导解决问题。求得 $f'(p) = C_{10}^8 2p^7 (1-p)(4-5p)$,则易知 $p=0.8$ 时, $f(p)$ 取得最大值。

第(Ⅱ)问求 p_k 的最大值,即求数列 $\{p_k\}$ 的最大项,即研究 $\{p_k\}$ 的单调性. 令 $\frac{p_{k+1}}{p_k} = \frac{C_{25}^{k+1} p^{k+1} (1-p)^{24-k}}{C_{25}^k p^k (1-p)^{25-k}} = \frac{(25-k)p}{(k+1)(1-p)} > 1$, 解得 $k < 26p - 1 = 19.8$. 则当 $k < 20$ 时, $p_{k+1} > p_k$, p_k 单调递增; 当 $k > 20$ 时, $p_{k+1} < p_k$, p_k 单调递减. 所以当 $k = 20$ 时, p_k 取得最大值.

教师追问: 设随机变量 $X \sim B(n, p)$, 记 $p_k = P(X = k)$, 求 p_k 取得最大值时 k 的取值.

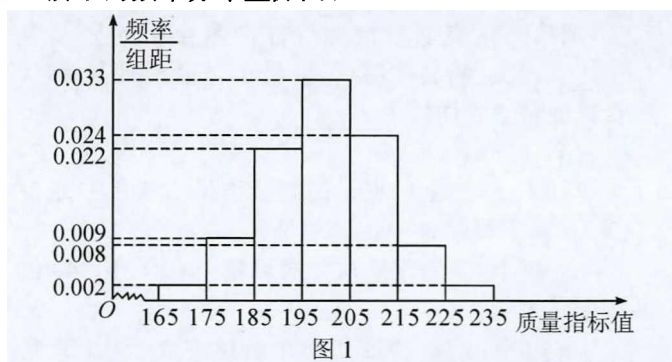
师生活动: 学生查阅人教 A 版《数学》(选择性必修三)第 81 页“探究与发现”, 仿照其探究过程, 自行推导结论. 可知, 当 $k < (n+1)p$ 时, $p_k > p_{k-1}$, p_k 随 k 值的增加而增加; 当 $k > (n+1)p$ 时, $p_k < p_{k-1}$, p_k 随 k 值的增加而减小. 若 $(n+1)p$ 为正整数, 当 $k = (n+1)p$ 时, $p_k = p_{k-1}$, 此时这两项概率均为最大值. 若 $(n+1)p$ 为非整数, 则取 $k = [(n+1)p]$, 此时 p_k 是唯一的最大值.

设计意图: 本题的两问体现了中学阶段求最值的两种主要手段: 一种是第(Ⅰ)问中的求导, 另外一种则是第(Ⅱ)问中的以数列观点解不等式. 由该题可以提炼出二项分布中概率最大项的一个结论, 为后续学生遇到同类问题时能快速解决打下基础.

3.2 理解正态分布的特征, 解决实际问题

正态分布问题主要考查学生能结合具体事例, 会用面积表示概率 $P(a \leq X \leq b)$, 会计算正态分布的概率; 能识别正态分布的参数对密度曲线的影响, 理解正态分布均值、方差及其含义, 能利用 3σ 原则解释策略的合理性.

例 2 从某企业生产的某种产品中抽取 500 件, 测量这些产品的一项质量指标值, 由测量结果得如图 1 所示的频率分布直方图:



由直方图可以认为, 这种产品的质量指标 Z 服从正态分布 $N(\mu, \sigma^2)$, 其中 μ 近似为样本平均数 $\bar{x}$, σ^2 近似为样本方差 s^2 .

(Ⅰ) 利用该正态分布, 求 $P(187.8 < Z < 212.2)$.

(Ⅱ) 某用户从该企业购买了 100 件这种产品, 记 X 表示这 100 件产品中质量指标值位于区间 $(187.8,$

$212.2)$ 的产品件数. 利用(Ⅰ)的结果, 求 $E(X)$.

附: $\sqrt{150} \approx 12.2$. 若 $Z \sim N(\mu, \sigma^2)$, 则 $P(\mu - \sigma < Z < \mu + \sigma) = 0.6826$, $P(\mu - 2\sigma < Z < \mu + 2\sigma) = 0.9544$.

师生活动: 题目要求利用频率分布直方图求出数据的均值和方差, 再用得到的均值和方差代替正态分布中的均值和方差. 观察频率分布直方图, 那么我们可以说这批零件的质量指标服从正态分布? 正态分布的概念是怎样的?

(Ⅰ) 由已知得, $Z \sim N(200, 150)$, 故 $P(187.8 < Z < 212.2) = P(200 - 12.2 < Z < 200 + 12.2) = 0.6826$.

师生活动: 该用户所购买的 100 件产品之间是否存在相互影响的关系? 每件产品的质量指标值位于区间 $(187.8, 212.2)$ 内的概率是否相同? 概率是多少? 应该怎样计算? 我们可以用什么样的概率模型来描述这一随机事件? 二项分布的数字特征又是怎样的?

(Ⅱ) 分析题意, 购买这 100 件产品, 即为 $X \sim B(100, 0.6826)$, 故 $E(X) = 100 \times 0.6826 = 68.26$.

设计意图: 一般来说, 若某一数量指标受很多随机因素的影响, 而每个单一因素的影响非常微小, 则这个指标服从正态分布. 正态分布有许多优良的性质, 许多分布可用正态分布来近似, 如二项分布、超几何分布等. 因此本题第(Ⅰ)问中用正态分布来解释测量的结果. 在第(Ⅱ)问中, 问题遵循 3σ 原则, 故将问题转化为二项分布做进一步处理. 解决此类问题需把握问题的本质, 建立合适的概率模型.

变式 2: 为了监控某种零件的一条生产线的生产过程, 检验员每天从该生产线上随机抽取 16 个零件, 并测量其尺寸(单位: cm). 根据长期生产经验, 可以认为这条生产线正常状态下生产的零件的尺寸服从正态分布 $N(\mu, \sigma^2)$.

(Ⅰ) 假设生产状态正常, 记 X 表示一天内抽取的 16 个零件中其尺寸在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之外的零件数, 求 $P(X \geq 1)$ 及 X 的数学期望.

(Ⅱ) 一天内抽检零件中, 如果出现了尺寸在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之外的零件, 就认为这条生产线在这一天的生产过程可能出现了异常情况, 需对当天的生产过程进行检查.

(i) 试说明上述监控生产过程方法的合理性;

(ii) 下面是检验员在一天内抽取的 16 个零件的尺寸(单位: cm):

9.95	10.12	9.96	9.96	10.01	9.92	9.98	10.04
10.26	9.91	10.13	10.02	9.22	10.04	10.05	9.95

经计算得 $\bar{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 9.97, s =$

$$\sqrt{\frac{1}{16} \sum_{i=1}^{16} (x_i - \bar{x})^2} = \sqrt{\frac{1}{16} (\sum_{i=1}^{16} x_i^2 - 16\bar{x}^2)} \approx 0.212,$$

其中 x_i 为抽取的第 i 个零件的尺寸, $i = 1, 2, \dots, 16$ 。用样本平均数 $\bar{x}$ 作为 μ 的估计值 $\hat{\mu}$, 用样本标准差 s 作为 σ 的估计值 $\hat{\sigma}$, 利用估计值判断是否需对当天的生产过程进行检查, 剔除 $(\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma})$ 之外的数据, 用剩下的数据估计 μ 和 σ (精确到 0.01)。

附: 若随机变量 Z 服从正态分布 $N(\mu, \sigma^2)$, 则 $P(\mu - 3\sigma < Z < \mu + 3\sigma) = 0.9974, 0.9974^{16} = 0.9592, \sqrt{0.008} \approx 0.09$ 。

解析: (I) 抽取的一个零件的尺寸在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之外的概率为 0.0026, 从而零件的尺寸在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之外的概率为 0.0026, 故 $X \sim B(16, 0.0026)$ 。因为 $P(X=0) \approx 0.9592$, 所以 $P(X \geq 1) = 1 - P(X=0) = 0.0408$ 。X 的数学期望为 $E(X) = 16 \times 0.0026 = 0.0416$ 。

(II) (i) 如果生产状态正常, 一个零件尺寸在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之外的概率只有 0.0026, 一天内抽取的 16 个零件中, 出现尺寸在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之外的零件的概率只有 0.0408, 发生的概率很小。因此一旦发生这种情况, 就有理由认为这条生产线在这一天的生产过程可能出现了异常情况, 需对当天的生产过程进行检查, 可见上述监控生产过程的方法是合理的。

(ii) 由 $\bar{x} = 9.97, s \approx 0.212$, 得 μ 的估计值为 $\hat{\mu} = 9.97, \sigma$ 的估计值为 $\hat{\sigma} = 0.212$, 由样本数据可以看出有一个零件的尺寸在 $(\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma})$ 之外, 因此需对当天的生产过程进行检查。

剔除 $(\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma})$ 之外的数据 9.22, 剩下数据的平均数为 $\frac{1}{15} \times (16 \times 9.97 - 9.22) = 10.02$, 因此 μ 的估计值为 10.02。 $\sum_{i=1}^{16} x_i^2 = 16 \times 0.212^2 + 16 \times 9.97^2 \approx 1591.134$, 剔除 $(\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma})$ 之外的数据 9.22, 剩下数据的样本方差为 $\frac{1}{15} \times (1591.134 - 9.22^2 - 15 \times 10.02^2) \approx 0.008$, 因此 σ 的估计值为 $\sqrt{0.008} \approx 0.09$ 。

设计意图: 解决正态分布问题, 首先要知道从频率分布直方图过渡到分布密度曲线的过程, 把握密度曲线的特征, 了解 3σ 原则, 即正态总体几乎总取值在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之内, 而在此之外取值的概率很小, 几乎不可能发生。若随机变量 X 服从正态分布 $N(\mu, \sigma^2)$, 则指标值落在 $(\mu - 3\sigma, \mu + 3\sigma)$ 的个数 Y 服从二项分布 $Y \sim B(n, 0.9974)$, 其他两个区间也类似。最后运用概率思想进行运算求解。

4 教学反思

4.1 认真研读教材, 精选复习素材

笔者建议教师在高三第二轮复习时不要提高所选例题的难度, 要学习课程标准中的“学业要求”, 认真研读教材, 理解二项分布、正态分布概念的内涵; 对于教材中的典型例题与习题, 体会其中蕴含的概率统计思想, 并适度挖掘, 选择具有含金量的试题作为复习的素材; 教师要在设计问题情境上下功夫, 选编的素材必须真实, 同时要让学生通过阅读, 经历抽象概括随机试验特征的过程, 从而理解两种分布模型的本质特征, 培养学生由生活实际到建立理论模型的推理能力。

4.2 关注生活实际, 提升建模能力

近几年的高考以社会生活、科学技术方面的素材作为试题情境, 条件中的数据经常以各种统计图表的形式出现, 加大了对学生阅读能力的考查。高三第二轮复习应强化学生审题训练, 结合条件中的关键词句, 给予学生数学建模的方法指导; 通过解决丰富多样、熟悉有趣的生活问题, 引导学生经历具体情境的分析到数学本质特征的抽象再到两种分布模型的建立的思维过程, 使学生体会两种分布模型的广泛应用及数学价值, 提升学生的数学建模、数据分析、数学运算等核心素养。

5 反馈训练

(1) 某陶瓷厂准备烧制甲、乙、丙三件不同的工艺品, 制作过程必须经过两次烧制, 当第一次烧制合格后方可进入第二次烧制, 两次烧制过程相互独立。根据该厂现有的技术水平, 第一次烧制, 甲、乙、丙三件产品合格的概率分别为 0.5, 0.6, 0.4, 第二次烧制, 甲、乙、丙三件产品合格的概率分别为 0.6, 0.5, 0.75, 则第一次烧制后恰有一件产品合格的概率为 _____; 经过两次烧制后, 合格工艺品的件数为 ξ , 则随机变量 ξ 的均值为 _____。

答案: 0.38, 0.9。

(2) (多选题) 某物理量的测量结果服从正态分布 $N(10, \sigma^2)$, 下列结论中正确的是()。

A. σ 越小, 该物理量在一次测量中在 $(9.9, 10.1)$ 的概率越大

B. σ 越小, 该物理量在一次测量中大于 10 的概率为 0.5

C. σ 越小, 该物理量在一次测量中小于 9.99 与大于 10.01 的概率相等

D. σ 越小, 该物理量在一次测量中落在 $(9.9, 10.2)$ 与落在 $(10, 10.3)$ 的概率相等

答案: ABC。