巧妙利用统计与概率的关系提升学生数学核心素养

——以一堂“超几何分布”课程设计为例

200231 上海市上海中学 刘 琴 刘 姗

摘 要:超几何分布是传统概率的一部分,是学生和教师都认识却不熟悉的分布,超几何分布的分布列研究和期望的证明一直是教与学的重点和难点.笔者通过一节公开课的教学设计,从实际背景出发,利用超几何分布和二项分布的对比,提出超几何分布的分布列讨论;利用现代科技手段探究超几何分布的期望;从统计的角度加强学生对组合数的理解,并指导学生完成超几何分布期望的严格证明;巧妙利用统计与概率的关系提升学生数学核心素养.

关键词:统计;概率;数学核心素养;超几何分布

学科核心素养是育人价值的集中体现,是学生通过学科学习而逐步形成的正确价值观、必备品格和关键能力.《普通高中数学课程标准》指出,数学核心素养包括数学抽象、逻辑推理、数学建模、直观想象、数学运算和数据分析.

面对最基础的数学课堂教学,教师一直在探索如何在课程设计中体现数学核心素养,让学生通过课程的学习学会在情境中抽象出数学概念,学会利用现代科技模拟数学问题,提升对数学问题的理解,积累依托数据探索事物本质、关联和规律的活动经验,形成规范化思考问题的品质,养成严谨求实的科学精神.

笔者以一节超几何分布课的课程设计为例,探讨数学核心素养在课堂中的体现.

一、情境问题

用熟悉的情境引入超几何分布这个学生相识却不相知的分布.

某商场为了吸引更多顾客,特在“双11”时举行抽奖活动,顾客从装有10只球(4只黄球、6只白球)的箱子里随机抽取2只球,若都是白球,则不中奖;若有1只黄球,则中二等奖,奖励购物券100元;若有2只黄球,则中一等奖,奖励购物券200元.每位顾客只能参与一次抽奖,商场向顾客提供以下两种选择.

1.采用放回抽样的方式,即每次取球后放回,充分混合后再抽取第二次.

2.采用不放回抽样的方式,即每次取球后不放回,从剩余的球中再抽取第二次.

问题1 如果你当天恰好在现场,你会采用哪种抽样方式?为什么?

问题2 如果你是商场总经理,你希望顾客采用哪种抽样方式?为什么?

设计意图:从实际情境出发,引导学生用数学语言表达实际问题,并以单元的视角提出研究超几何分布的必要性,激发学生探索的兴趣.

简化问题 已知箱中有10只球(4只黄球、6只白球),从中随机抽取2只.

1.若每次抽取后放回,设抽到黄球的个数为X,求X的分布列.

$P(X=0)=C_2^0\cdot0.4^0\cdot0.6^2$

$P(X=1)=C_2^1\cdot0.4^1\cdot0.6^1$

$P(X=2)=C_2^2\cdot0.4^2\cdot0.6^0$

X	0	1	2
$P(X)$	0.36	0.48	0.16

$E(X)=0\cdot0.36+1\cdot0.48+2\cdot0.16=0.8$.

2.若每次抽取后不放回,设抽到的黄球个数为Y,求Y的分布列.

$$P(Y=0)=\frac{C_4^0\cdot C_6^2}{C_{10}^2}$$

$$P(Y=1)=\frac{C_4^1\cdot C_6^1}{C_{10}^2}$$

$$P(Y=2)=\frac{C_4^2\cdot C_6^0}{C_{10}^2}$$

Y	0	1	2
$P(Y)$	$0.3\dot{3}$	$0.5\dot{3}$	$0.1\dot{3}$

$E(Y)=0\cdot0.3\dot{3}+1\cdot0.5\dot{3}+2\cdot0.1\dot{3}=0.8$.

思考1 从顾客的角度来看,为什么选择放回抽样?为什么选择不放回抽样?(从决策论的角度探讨)

引导 对比获得不同奖励的概率,如获得奖励的概率或获得大奖的概率.

思考2 两个分布的期望是一样的,这是巧合还是存在某种内在的联系?

引导 放回抽样是学生已经熟悉的二项分布,带领学生回忆二项分布的条件.

提问 不放回抽样的分布是不是也有一定的规律?

变式1 若袋中有10只球(4只黄球、6只白球),从中随机抽取5次,都不放回,设抽到的黄球个数为Y,求Y的分布列.

设计意图:在不放回抽样的情形下,黄球个数Y的取值会受到取球个数和袋中原有黄球个数的限制,故设计此变式,引导学生讨论Y的取值,为进一步抽象概括做准备.引导学生讨论如下.

1.抽到的黄球个数Y可能的取值是多少?(引导学生讨论Y的最大可能取值是在原有黄球个数和抽取球个数中取大)

2.对应不同取值的概率是多少?(引导学生在数字改变的情况下进一步探索超几何分布的分布列)

$$P(Y=0)=\frac{C_4^0\cdot C_6^5}{C_{10}^5}$$

$$P(Y=1)=\frac{C_4^1\cdot C_6^4}{C_{10}^5}$$

$$P(Y=2)=\frac{C_4^2\cdot C_6^3}{C_{10}^5}$$

$$P(Y=3)=\frac{C_4^3\cdot C_6^2}{C_{10}^5}$$

$$P(Y=4)=\frac{C_4^4\cdot C_6^1}{C_{10}^5}$$

所以 $P(Y=k)=\frac{C_4^k\cdot C_6^{5-k}}{C_{10}^5},k=0,1,\cdots,4.$

二、抽象概括

问题3 如果将实际问题中的具体数据变成字母,是否可以抽象出超几何分布的分布列?

变式2 若袋中有N只球[M只黄球、$(N-M)$只白球],从中随机抽取n次,都不放回,设抽到的黄球个数为Y,求Y的分布列.

设计意图:从具体的数字到抽象的表达是形成理性思维的过程,利用熟悉的情境进行抽象概括,为更一般情境下的抽象定义奠定基础.

问题4 如果不以球作为背景,能否进一步抽象出超几何分布的定义?

引导 讨论并提出超几何分布的定义.

超几何分布 设N个元素分为两类,其中M个属于第一类,$(N-M)$个属于第二类.从中按不重复抽样取n个,令Y表示这n个样品中第一类的元素个数,则Y的分布称为超几何分布(Hypergeometric Distribution).记作$Y\sim H(n,M,N)$,且$P(Y=k)=\frac{C_M^k\cdot C_{N-M}^{n-k}}{C_N^n},k=0,1,2\cdots,m,m=\min(M,n)$,$n\leqslant N,M\leqslant N,n,M,N$为正整数.

Y的分布列如表1所示.

三、归纳猜想

问题5 我们在情境问题中讨论过,顾客会根据自己对获得奖励与否或奖励金额的偏好进行选择,但从商场经理的角度而言,两种选择所需发放的购物券金额是一样的,这一结论由两种分布的期望所支持.如果有放回地抽样,X表示抽到的黄球个数,则X服从二项分布$B(n,p)$,期望是$E(X)=np$.如果不放回地抽样,Y表示抽到的黄球个数,则Y服从超几何分布,超几何分布的期望是什么?

设计意图:鼓励学生大胆猜想,试图找到超几何分布与二项分布的期望的相关性.

四、模拟实验

问题6 为了验证我们的猜想是否可行,在试图证明之前,我们还可以用什么样的方式进行探究?

设计意图:鼓励学生用不同方式进行探究,验证猜想$E(Y)=n\cdot\frac{M}{N}$成立的可能性.可以是实物模拟(从袋中取球),也可以是用图形计算器生成随机数进行模拟(具体操作如图1所示),更鼓励学生使用大数据模拟的方式(学生编写了R语言程序,具体操作如图2所示,具体链接为https://charlieji－r.shinyapps.io/Hypergeometric_Simulation_APP_CharlieJi/).

表1

Y	0	1	2	…	$m-1$	m
$P(Y)$	$\frac{C_M^0\cdot C_{N-M}^n}{C_N^n}$	$\frac{C_M^1\cdot C_{N-M}^{n-1}}{C_N^n}$	$\frac{C_M^2\cdot C_{N-M}^{n-2}}{C_N^n}$	…	$\frac{C_M^{m-1}\cdot C_{N-M}^{n-m+1}}{C_N^n}$	$\frac{C_M^m\cdot C_{N-M}^{n-m}}{C_N^n}$

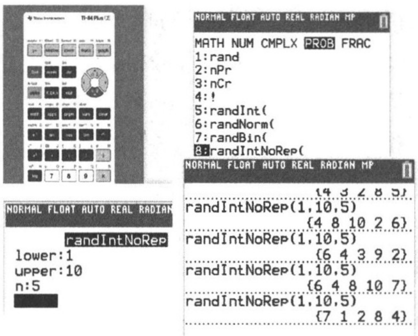

图1　图形计算器生成随机数模拟抽样

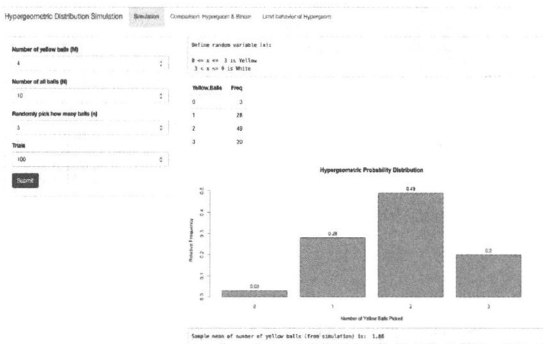

图2　R语言程序生成随机数模拟抽样

五、推理证明

问题7　经过模拟探究后,应该进行严格的推理证明.

设计意图:对猜想的严格证明可通过组合数的展开进行,但计算比较繁琐,若利用组合数的性质,一方面可以使证明过程更加简洁,另一方面可以加深学生对组合数性质的理解.

引理1　$kC_M^k=MC_{M-1}^{k-1}$.

引理2　$C_M^0C_{N-M}^n+C_M^1C_{N-M}^{n-1}+C_M^2C_{N-M}^{n-2}+\cdots+C_M^mC_{N-M}^{n-m}=C_N^n$.

第一步:探究引理1

严格证明:$\because C_N^n=\dfrac{N!}{n!(N-n)!}$,

$\therefore kC_M^k=k\cdot\dfrac{M!}{k!(M-k)!}=\dfrac{M!}{(k-1)!(M-k)!}$

$=M\cdot\dfrac{(M-1)!}{(k-1)!(M-k)!}=M\cdot C_{M-1}^{k-1}$.

概率的思想理解如图3所示.

第二步:证明$E(Y)=n\cdot\dfrac{M}{N}$

因为如表1所示,由期望的定义得$E(Y)=0\cdot\dfrac{C_M^0C_{N-M}^n}{C_N^n}+1\cdot\dfrac{C_M^1C_{N-M}^{n-1}}{C_N^n}+2\cdot\dfrac{C_M^2C_{N-M}^{n-2}}{C_N^n}+\cdots+m\cdot\dfrac{C_M^mC_{N-M}^{n-m}}{C_N^n}$.

由引理1,$kC_M^k=MC_{M-1}^{k-1}$,将所有分子的前半部分乘法简化如下:

$$E(Y)=0\cdot\frac{C_M^0C_{N-M}^n}{C_N^n}+1\cdot\frac{C_M^1C_{N-M}^{n-1}}{C_N^n}+2\cdot\frac{C_M^2C_{N-M}^{n-2}}{C_N^n}+\cdots+m\cdot\frac{C_M^mC_{N-M}^{n-m}}{C_N^n}=\frac{1}{C_N^n}(1\cdot C_M^1C_{N-M}^{n-1}+2\cdot C_M^2C_{N-M}^{n-2}+\cdots+m\cdot C_M^mC_{N-M}^{n-m})=\frac{1}{C_N^n}(M\cdot C_{M-1}^0C_{N-M}^{n-1}+M\cdot C_{M-1}^1C_{N-M}^{n-2}+\cdots+M\cdot C_{M-1}^{m-1}C_{N-M}^{n-m})$$①.

进一步提出共同项,得到①式$=\dfrac{M}{C_N^n}(C_{M-1}^0C_{N-M}^{n-1}+C_{M-1}^1C_{N-M}^{n-2}+\cdots+C_{M-1}^{m-1}C_{N-M}^{n-m})$.

提问　此时括号中的和是多少?

引导　学生从分布列的性质讨论,得到引理2:$C_M^0C_{N-M}^n+C_M^1C_{N-M}^{n-1}+C_M^2C_{N-M}^{n-2}+\cdots+C_M^mC_{N-M}^{n-m}=C_N^n$.

思考　如表1,分布列中所有概率的和是多少?

学生已经了解分布列的性质,能够回答和为1,即$\dfrac{C_M^0C_{N-M}^n}{C_N^n}+\dfrac{C_M^1C_{N-M}^{n-1}}{C_N^n}+\dfrac{C_M^2C_{N-M}^{n-2}}{C_N^n}+\cdots+\dfrac{C_M^mC_{N-M}^{n-m}}{C_N^n}=1$.

将分母移到右边,得到$C_M^0C_{N-M}^n+C_M^1C_{N-M}^{n-1}+C_M^2C_{N-M}^{n-2}+\cdots+C_M^mC_{N-M}^{n-m}=C_N^n$.

思考　证明过程中,括号中的和是多少?

通过如图4所示的过程,学生能够快速了解要求的和等价于从$(N-1)$个元素中取出$(n-1)$个第一类元素的所有情况,故可以得到:

括号中的和$=\dfrac{M}{C_N^n}(C_{M-1}^0C_{N-M}^{n-1}+C_{M-1}^1C_{N-M}^{n-2}+\cdots$

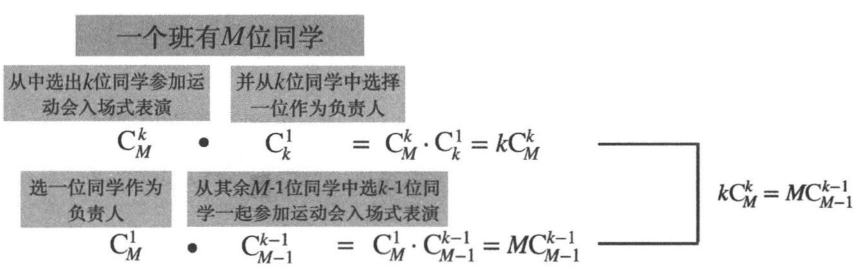

图3

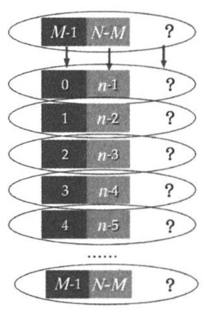

图4

$+C_{M-1}^{m-1}C_{N-M}^{n-m})=\frac{M}{C_N^n}\cdot C_{N-1}^{n-1}$ ②.

利用组合数的定义，学生能够快速完成余下的证明，过程如下：

$$②式=M\cdot\frac{(N-1)!}{(n-1)!\ (N-n)!}\cdot\frac{n!\ (N-n)!}{N!}$$

$$=M\cdot\frac{1}{(n-1)!}\cdot\frac{n!}{N}=M\cdot\frac{n}{N}.$$

六、总结归纳

(一)超几何分布与二项分布的关系

超几何分布与二项分布的关系如图5-1所示.

1. 当N足够大时，放回和不放回对概率的影响很小，超几何分布逼近二项分布.

2. 利用R语言探究二者关系(如图5-2所示，具体链接为https://charlieji－r.shinyapps.io/Hypergeometric_Simulation_APP_CharlieJi/).

(二)探索方法总结

本节课的探索方法为“观察—猜想—模拟—大数据模拟—严格证明”，这一探索方法(过程)非常重要.

七、结语

本节课是一节“双新”展示公开课，可以在“研直播”观看(研直播—教研活动—数学学科教学展示课及专家点评)，本节课的设计突出以下方面.

首先，从单元的角度来看，从实际情境出发，结合二项分布提出超几何分布的概念及其均值的探索与研究，培养学生从情境中抽象出数学概念、命题、方法和体系的能力. 其次，本节课大胆利用统计与概率的密切联系，从统计的角度鼓励学生进行超几何分布的研究，并利用统计的思想“一两拨千金”地完成规范化的证明，学生形成规范化思考问题的品质，养成严谨求实的科学精神. 从教学过程来看，本节课非常注重数字化学习的引导，利用图形计算器、R语言等工具模拟随机过程，帮助学生探索；注重统计课程的完整性，引导学生完整地经历了观察、猜想、模拟、证明的过程.

参考文献

[1] David E. Bock, Paul F. Velleman, Richard D. De Veaux. Stats Modeling the World: 3rd Edition [M]. New York: Addison-Wesley, 2008: 388-399.

[2] 鲍毅. 关于超几何分布的探究[J]. 考试周刊, 2017(77): 69.

[3] 人民教育出版社 课程教材研究所，中学数学课程教材研究开发中心. 普通高中教科书·数学：选择性必修第二册B版[M]. 北京：人民教育出版社, 2019: 75-79.

[4] 郭庆平. 公式 $kC_n^k=nC_{n-1}^{k-1}$ 与组长选举[J]. 中学数学, 1988(8): 30.

[5] 赵永祥，李丽. 超几何概率教学探讨[J]. 重庆三峡学院学报, 2014(3): 32-34.

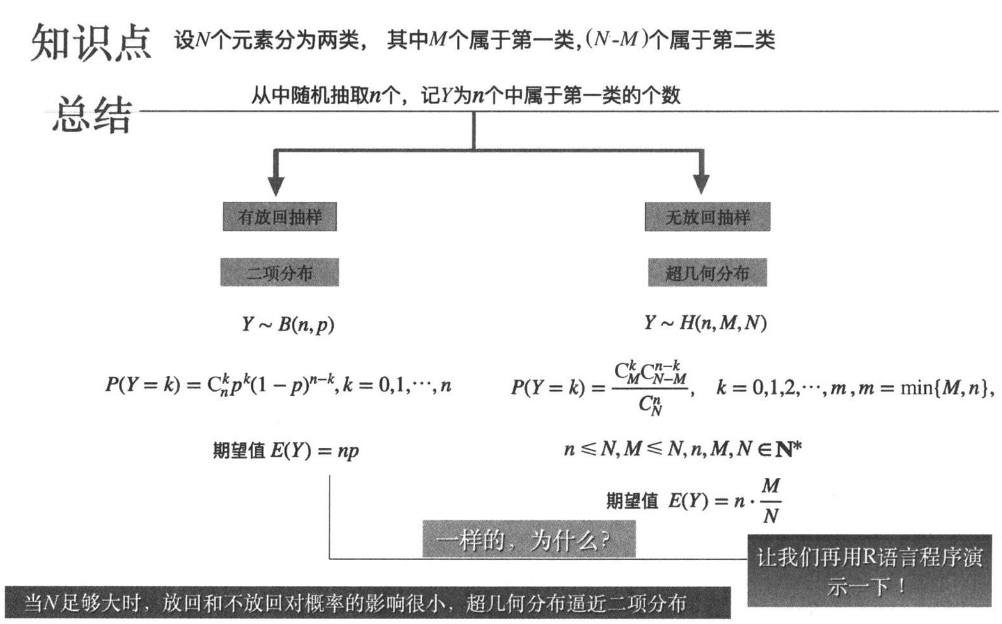

图5-1 二项分布与超几何分布的知识点总结

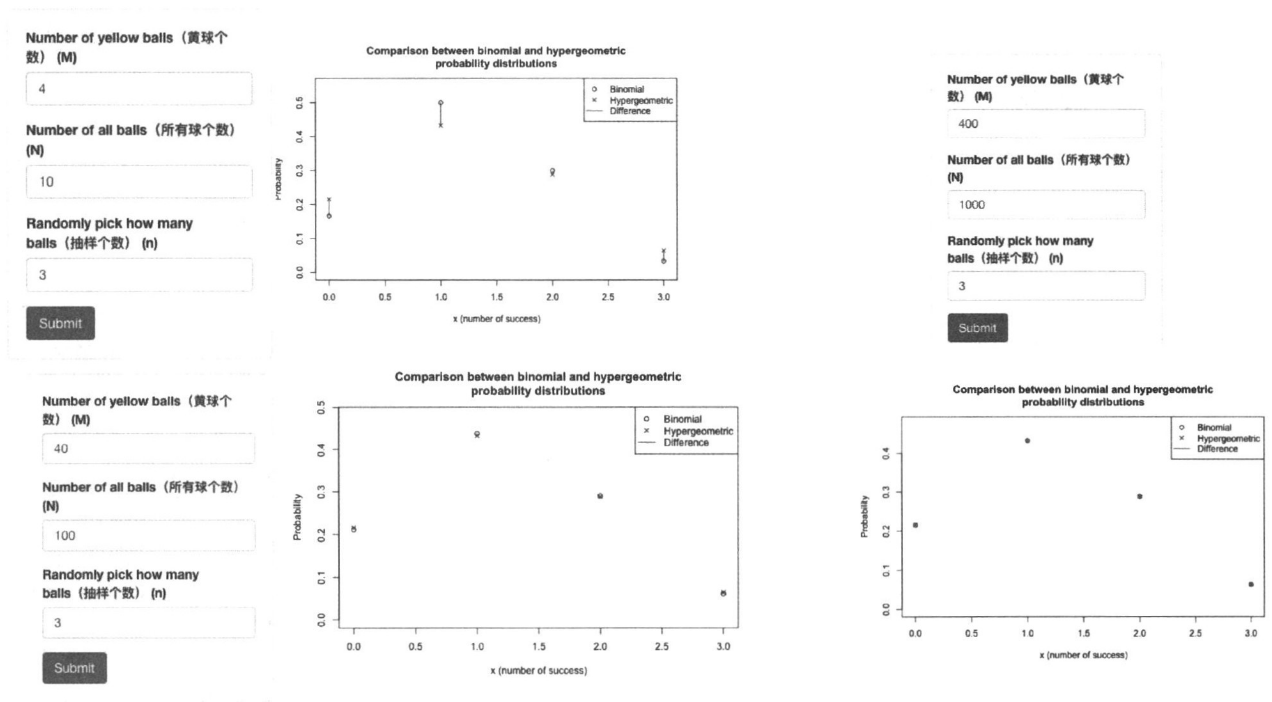

图 5-2　随着 N 的不同，超几何分布与二项分布的取值比较

（上接第 30 页）

$=a^2+2ab+b^2$ 以及 $(a+b)(a-b)=a^2-b^2$，此时就可以参考浙教版教材“如果把 $(a-b)^2$ 写成 $[a+(-b)]^2$，就可以由两数和的完全平方公式写出两数差的完全平方公式”，将 $(a+b)^2=a^2+2ab+b^2$ 与 $(a-b)^2=a^2-2ab+b^2$ 两个公式进行比较.

这样的引领性问题体现了如图 6 所示的乘法公式的引出过程，更多地体现出学生的自主探究.

公式探究 ⇨ 公式归纳 ⇨ 文字表述 ⇨ 差异比较

图 6

以上两个方案在探究过程中，都关注到乘法公式的特殊性，从数和形两个角度进行研究，在保持研究角度一致的情况下，需要从数的角度对从形的角度观察得到的结果进行验证，也可以从形的角度对从数的角度探究归纳得到的结果进行说明与表示. 因此，这两种方案设计都可以在一个课时的学生探究过程中得出三个特殊的乘法公式，同时也达成了分开推导所不能达到的目标：分辨各个公式各有什么样的特点，以及进行各个公式差异的比较.

知识结构化不仅体现在单元或章节的复习环节中，而且应该体现在新课的教学中. 找到知识的生长点，从结构生成的角度进行合理、有效的设计，并且设计好问题与活动，重点体现学生对于新问题的研究思路与探究过程，这就是单元教学的魅力与价值所在.

参考文献

[1] 章建跃. 数学课堂教学设计研究[J]. 数学通报，2006(7)：20-26.

[2] 徐淀芳，刘达. 初中数学单元教学设计指南[M]. 北京：人民教育出版社，2018.

[3] 上海市中小学(幼儿园)课程改革委员会. 九年义务教育课本·数学：七年级第一学期试用本[M]. 上海：上海教育出版社，2019.

[4] 上海市中小学(幼儿园)课程改革委员会. 九年义务教育数学教学参考资料：七年级第一学期试用本[M]. 上海：上海教育出版社，2019.

[5] 人民教育出版社 课程教材研究所 中学数学课程教材研究开发中心. 义务教育教科书·数学：八年级上册[M]. 北京：人民教育出版社，2013.

[6] 马复. 义务教育教科书·数学：七年级下册[M]. 北京：北京师范大学出版社，2012.

[7] 范良火. 义务教育教科书·数学：七年级下册[M]. 杭州：浙江教育出版社，2012.