（2014-2023）十年概率统计高考题研究　　2021 级高三数学 张红青

题号	难度	知识点
2023·全国·甲卷理	适中	计算几个数的中位数,独立性检验解决实际问题,超几何分布的均值,超几何分布的分布列
2023·全国乙卷	较易	计算几个数的平均数,计算几个数据的极差、方差、标准差,统计新定义
2023·全国·1 卷	适中	求离散型随机变量的均值,利用全概率公式求概率
2023·全国·2 卷	适中	频率分布直方图的实际应用,总体百分位数的估计
2023·全国·甲卷文	适中	计算几个数的平均数,完善列联表,卡方的计算，独立性检验解决实际问题
2022·全国 2 卷	适中	频率分布直方图的实际应用,由频率分布直方图估计平均数,利用对立事件的概率公式求概率,计算条件概率
2022·全国·甲卷文	容易	卡方的计算,计算古典概型问题的概率
2022·全国·甲卷理	较易	写出简单离散型随机变量分布列,求离散型随机变量的均值
2022·全国·乙卷	适中	计算几个数的平均数,相关系数的计算,根据样本中心点求参数
2022·全国 1 卷	适中	独立性检验解决实际问题,计算条件概率
2021·全国·2 卷	较难	利用导数研究方程的根,求离散型随机变量的均值,均值的实际应用
2021·全国·甲卷	较易	独立性检验解决实际问题
2021·全国·乙卷	容易	计算几个数的平均数,计算几个数据的极差、方差、标准差
2021·全国 1 卷	较易	写出简单离散型随机变量分布列,求离散型随机变量的均值
2020·全国·3 卷	适中	独立性检验解决实际问题
2020·全国·1 卷文	适中	用平均数的代表意义解决实际问题,计算频率
2020·全国 1 卷理	适中	独立事件的实际应用
2020·全国·2 卷	适中	相关系数的计算

2019 全国 1 卷理 21	较难	马尔科夫链，与数列知识结合，递推公式的应用
2019 北京 17	较易	离散型随机变量分布列,求离散型随机变量的均值，小概率事件
2019 全国 3 卷 17	较易	频率分布直方图，计算平均值
2019 全国 2 卷 18	适中	比赛类 独立事件的实际应用，古典概型的概率
2019 天津 16	较易	二项分布、独立事件的实际应用
2019 江苏 25	适中	列举法，古典概型问题的概率，对立事件的概率公式求概率
2018 课标 1 卷 20	较难	产品合格不合格问题，利用导数求最值点，二项分布、决策问题
2018 课标 2 卷 18	适中	折线图、回归模型拟合效果分析
2018 课标 3 卷 18	适中	茎叶图、独立性检验
2018 北京 17	较易	独立事件的实际应用，分布列、期望、方差，二点分布，决策
2018 天津 16	较易	分层抽样、互斥事件、超几何分布的分布列、期望
2017 课标 1 卷 19	适中	正态分布、产品检测，剔除数据以后新样本的平均数、方差
2017 课标 2 卷 18	适中	频率分布直方图、计算中位数，填写 2*2 列联表，独立性检验
2017 课标 3 卷 18	适中	超市利润决策，分布列、期望、方差
2017 北京 17	较易	图表分析方差，超几何分布的分布列、期望
2017 天津 16	较易	信号灯，独立事件，分布列、期望
2017 山东 18	适中	超几何分布的分布列，期望
2017 江苏 26	较难	摸球放入抽屉，分布列，期望，证明不等式（放缩后使用组合数的性质求和）
2016 课标 1 卷 19	适中	柱状图，分布列，期望，决策
2016 课标 2 卷 18	适中	保险，互斥事件概率加法，条件概率，分布列、期望
2016 课标 3 卷 18	适中	折线图、（生活垃圾无公害化处理），相关系数，回归方程

2016 天津 16	较易	离散型随机变量的分布列、期望
2016 山东 19	适中	比赛类，独立事件，互斥事件概率加法，分布列、期望
2016 北京 16	较易	分层抽样、互斥事件、独立事件、古典概型、平均数
2016 四川 16	较易	频率分布直方图、样本估计总体，85%分位数
2015 课标 1 卷 19	较难	散点图、非线性回归，回归方程
2015 课标 2 卷 18	适中	茎叶图、平均值、方差、独立事件
2015 广东 17	适中	系统抽样、均值、方差
2015 重庆 17	适中	超几何分布、古典概型的分布列、期望
2015 天津 16	适中	（乒乓球）超几何分布，分布列、期望
2015 四川 17	适中	超几何分布、对立事件、分布列、期望
2015 陕西 19	适中	对立事件求概率，独立事件，分布列，期望
2015 山东 19	适中	数字问题，得分分布列，期望
2015 湖南 20	适中	（摸球）互斥事件、独立事件，二项分布，分布列、期望
2015 湖北 20	适中	（利润）分布列期望，独立事件，二项分布，线性规划求最值
2015 福建 16	较易	（银行卡密码）数字问题，分布列、期望
2015 北京 16	较易	列举法，古典概型，比较两组数据的方差
2015 安徽 17	适中	（产品正品次品检测）不放回，分布列、期望
2014 课标 1 卷 18	适中	频率分布直方图，平均数，方差，正态分布
2014 课标 2 卷 19	适中	线性回归方程
2014 大纲 20	较易	独立事件
2014 北京 16	较易	（篮球比赛）独立事件

2014 安徽 17	**适中**	**（围棋比赛）独立事件，分布列，期望**
2014 广东 17	**适中**	**频率分布表、由样本估计总体，二项分布**
2014 陕西 21	**适中**	**(农作物种植利润)独立事件**
2014 重庆 18	**适中**	**数字问题，古典概型分布列**
2014 天津 16	**较易**	**超几何分布、分布列、期望**
2014 四川 17	**适中**	**（游戏得分）独立事件，分布列，期望**
2014 山东 18	**较难**	**（乒乓球比赛）独立事件，分布列，期望**
2014 辽宁 18	**适中**	**频率分布直方图，独立事件，二项分布，分布列、期望**
2014 江西 22	**较难**	**数学归纳法，分布列，期望，对立事件**
2014 湖南 17	**适中**	**独立事件，分布列，期望**
2014 湖北 20	**适中**	**（利润决策问题）独立事件，二项分布，期望**
2014 江苏 25	**适中**	**（抽球）分布列，期望**
2014 福建 18	**适中**	**（商场兑奖决策问题）分布列，期望，方差**

难题展示：

1.（2019·全国Ⅰ·理·第 21 题）为治疗某种疾病，研制了甲、乙两种新药，希望知道哪种新药更有效，为此进行动物试验．试验方案如下：每一轮选取两只白鼠对药效进行对比试验．对于两只白鼠，随机选一只施以甲药，另一只施以乙药．一轮的治疗结果得出后，再安排下一轮试验．当其中一种药治愈的白鼠比另一种药治愈的白鼠多 4 只时，就停止试验，并认为治愈只数多的药更有效．为了方便描述问题，约定，对于每轮试验，若施以甲药的白鼠治愈且施以乙药的白鼠未治愈则甲药得 1 分，乙药得－1 分；若施以乙药的白鼠治愈且施以甲药的白鼠未治愈则乙药得 1 分，甲药得－1 分；若都治愈或都未治愈则两种药均得 0 分．甲、乙两种药的治愈率分别记为 α 和 β，一轮试验中甲药的得分记为 X.

(1) 求 X 的分布列；

(2) 若甲药、乙药在试验开始时都赋予 4 分，$p_i(i=0,1,\cdots,8)$ 表示“甲药的累计得分为 i 时，最终认为甲药比乙药更有效”的概率，则

$p_0=0,\ p_8=1,\ p_i=ap_{i-1}+bp_i+cp_{i+1}$（$i=1,2,\cdots,7$），

其中 $a=P(X=-1)$， $b=P(X=0)$， $c=P(X=1)$． 假设 $\alpha=0.5$， $\beta=0.8$．

(i) 证明： $\{p_{i+1}-p_i\}$ $(i=0,1,2,\cdots,7)$ 为等比数列；

(ii) 求 p_4， 并根据 p_4 的值解释这种试验方案的合理性．

【答案】(1) 解：X 的所有可能取值为 $-1,0,1$，

$P(X=-1)=(1-\alpha)\beta$，$P(X=0)=\alpha\beta+(1-\alpha)(1-\beta)$，$P(X=1)=\alpha(1-\beta)$．

所以 X 的分布列为

X	-1	0	1
P	$(1-\alpha)\beta$	$\alpha\beta+(1-\alpha)(1-\beta)$	$\alpha(1-\beta)$

(2) (i) 由 (1) 得 $a=0.4,\quad b=0.5,\quad c=0.1$．

因此 $p_i=0.4p_{i-1}+0.5p_i+0.1p_{i+1}$， 故 $0.1(p_{i+1}-p_i)=0.4(p_i-p_{i-1})$， 即 $p_{i+1}-p_i=4(p_i-p_{i-1})$．

又因为 $p_1-p_0=p_1\neq 0$， 所以 $\{p_{i+1}-p_i\}(i=0,1,2,\cdots,7)$ 为公比为 4， 首项为 p_1 的等比数列.

(ii) 由 (i) 可得

$$p_8=p_8-p_7+p_7-p_6+\cdots+p_1-p_0+p_0=(p_8-p_7)+(p_7-p_6)+\cdots+(p_1-p_0)=\frac{4^8-1}{3}p_1 .$$

由于 $p_8=1$， 故 $p_1=\dfrac{3}{4^8-1}$， 所以 $p_4=(p_4-p_3)+(p_3-p_2)+(p_2-p_1)+(p_1-p_0)=\dfrac{4^4-1}{3}p_1=\dfrac{1}{257}$．

p_4 表示最终认为甲药更有效的概率，由计算结果可以看出，在甲药治愈率为 0.5，乙药治愈率为 0.8 时，认为甲药更有效的概率为 $p_4=\dfrac{1}{257}\approx 0.0039$，此时得出错误结论的概率非常小，说明这种试验方案合理.

【点睛】本题考查离散型随机变量分布列的求解、利用递推关系式证明等比数列、累加法求解数列通项公式和数列中的项的问题.本题综合性较强，要求学生能够熟练掌握数列通项求解、概率求解的相关知识，对学生分析和解决问题能力要求较高.

2.(2018 年高考数学课标卷Ⅰ(理)·第 20 题)(12 分)某工厂的某种产品成箱包装，每箱 200 件，每一箱产品在交付用户之前要对产品作检验，如检验出不合格品，则更换为合格品．检验时，先从这箱产品中任取 20 件作检验，再根据检验结果决定是否对余下的所有产品作检验，设每件产品为不合格品的概率都为 $p(0<p<1)$，且各件产品是否为不合格品相互独立.

(1)记 20 件产品中恰有 2 件不合格品的概率为 $f(p)$,求 $f(p)$ 的最大值点 p_0.

(2)现对一箱产品检验了 20 件，结果恰有 2 件不合格品，以(1)中确定的 p_0 作为 p 的值．已知每件产品的检验费用为 2 元，若有不合格品进入用户手中，则工厂要对每件不合格品支付 25 元的赔偿费用.

(i)若不对该箱余下的产品作检验，这一箱产品的检验费用与赔偿费用的和记为 X，求 EX;

(ii)以检验费用与赔偿费用和的期望值为决策依据，是否该对这箱余下的所有产品作检验?

【答案】解析：(1)20 件产品中恰有 2 件不合格品的概率为 $f(p)=\mathrm{C}_{20}^{2}p^{2}(1-p)^{18}$.

因此 $f'(p)=\mathrm{C}_{20}^{2}[2p(1-p)^{18}-18p^{2}(1-p)^{17}]=2\mathrm{C}_{20}^{2}p(1-p)^{17}(1-10p)$.

令 $f'(p)=0$，得 $p=0.1$．当 $p\in(0,0.1)$时，$f'(p)>0$；当 $p\in(0.1,1)$时，$f'(p)<0$．所以 $f(p)$ 的最大值点为 $p_0=0.1$.

(2)由(1)知，$p=0.1$.

(i)令 Y 表示余下的 180 件产品中的不合格品件数，依题意知 $Y\sim B(180,0.1)$，$X=20\times2+25Y$，即 $X=40+25Y$.

所以 $EX=E(40+25Y)=40+25EY=490$.

(ii)如果对余下的产品作检验，则这一箱产品所需要的检验费为 400 元．由于 $EX>400$，故应该对余下的产品作检验.

3.(2017年高考数学江苏文理科•第26题)已知一个口袋有 m 个白球, n 个黑球($m,n\in\mathbf{N}^*,n\geqslant 2$),这些球除颜色外全部相同．现将口袋中的球随机的逐个取出,并放入如图所示的编号为 $1,2,3,\cdots,m+n$ 的抽屉内,其中第 k 次取出的球放入编号为 k 的抽屉 $(k=1,2,3,\cdots,m+n)$ ．

1	2	3	$\cdots$	$m+n$

(1)试求编号为2的抽屉内放的是黑球的概率 p ;

(2)随机变量 X 表示最后一个取出的黑球所在抽屉编号的倒数, $E(X)$ 是 X 的数学期望,证明: $E(X)<\dfrac{n}{(m+n)(n-1)}$

【答案】(1) $\dfrac{n}{m+n}$ (2)见解析

解析:解:(1)编号为2的抽屉内放的是黑球的概率 p 为: $p=\dfrac{\mathrm{C}_{m+n-1}^{n-1}}{\mathrm{C}_{m+n}^{n}}=\dfrac{n}{m+n}$.

(2)随机变量 X 的概率分布为:

X	$\dfrac{1}{n}$	$\dfrac{1}{n+1}$	$\dfrac{1}{n+2}$		$\dfrac{1}{k}$		$\dfrac{1}{m+n}$
P	$\dfrac{\mathrm{C}_{n-1}^{n-1}}{\mathrm{C}_{m+n}^{n}}$	$\dfrac{\mathrm{C}_{n}^{n-1}}{\mathrm{C}_{m+n}^{n}}$	$\dfrac{\mathrm{C}_{n+1}^{n-1}}{\mathrm{C}_{m+n}^{n}}$		$\dfrac{\mathrm{C}_{k-1}^{n-1}}{\mathrm{C}_{m+n}^{n}}$		$\dfrac{\mathrm{C}_{n+m-1}^{n-1}}{\mathrm{C}_{m+n}^{n}}$

随机变量 X 的期望为:

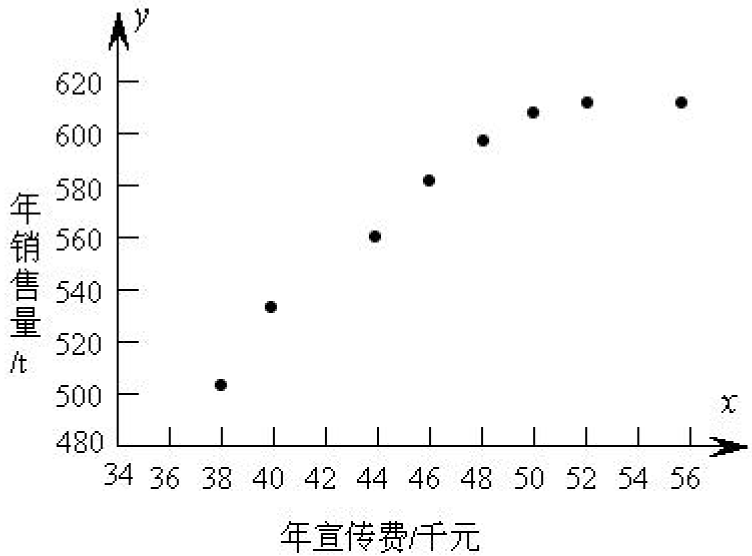
$$E(X)=\sum_{k=n}^{m+n}\frac{1}{k}\cdot\frac{C_{k-1}^{n-1}}{C_{m+n}^{n}}=\frac{1}{C_{m+n}^{n}}\sum_{k=n}^{m+n}\frac{1}{k}\cdot\frac{(k-1)!}{(n-1)!(k-n)!},$$

所以 $E(X)<\dfrac{1}{C_{m+n}^{n}}\displaystyle\sum_{k=n}^{m+n}\frac{(k-2)!}{(n-1)!(k-n)!}=\frac{1}{(n-1)C_{m+n}^{n}}\sum_{k=n}^{m+n}\frac{(k-2)!}{(n-2)!(k-n)!}=\frac{1}{(n-1)C_{m+n}^{n}}(1+C_{n-1}^{n-2}+C_{n}^{n-2}+\cdots+C_{m+n-2}^{n-2})$

$$=\frac{1}{(n-1)C_{m+n}^{n}}(C_{n-1}^{n-1}+C_{n-1}^{n-2}+C_{n}^{n-2}+\cdots+C_{m+n-2}^{n-2})=\frac{1}{(n-1)C_{m+n}^{n}}(C_{n}^{n-1}+C_{n}^{n-2}+\cdots+C_{m+n-2}^{n-2})=\cdots=\frac{1}{(n-1)C_{m+n}^{n}}(C_{m+n-2}^{n-1}+C_{m+n-2}^{n-2})$$

$=\dfrac{C_{m+n-1}^{n-1}}{(n-1)C_{m+n}^{n}}=\dfrac{n}{(m+n)(n-1)}$ 所以 $E(X)<\dfrac{n}{(m+n)(n-1)}$.

4.(2015 高考数学新课标 1 理科·第 19 题)某公司为确定下一年度投入某种产品的宣传费，需了解年宣传费 x(单位：千元)对年销售量 y(单位：t)和年利润 z(单位：千元)的影响，对近 8 年的年宣传费 x_i 和年销售量 y_i($i=1,2,\cdots,8$)数据作了初步处理，得到下面的散点图及一些统计量的值。



$\overline{x}$	$\overline{y}$	$\overline{w}$	$\sum\limits_{i=1}^{8}(x_i-\overline{x})^2$	$\sum\limits_{i=1}^{8}(w_i-\overline{w})^2$	$\sum\limits_{i=1}^{8}(x_i-\overline{x})(y_i-\overline{y})$	$\sum\limits_{i=1}^{8}(w_i-\overline{w})(y_i-\overline{y})$
46．6	56．3	6．8	289．8	1．6	1469	108．8

表中 $w_i=\sqrt{x_i}$，$\overline{w}=\sum\limits_{i=1}^{8}w_i$。

(Ⅰ)根据散点图判断，$y=a+bx$ 与 $y=c+d\sqrt{x}$ 哪一个适宜作为年销售量 y 关于年宣传费 x 的回归方程类型？(给出判断即可，不必说明理由)

(Ⅱ)根据(Ⅰ)的判断结果及表中数据，建立 y 关于 x 的回归方程；

(III) 已知这种产品的年利率 z 与 x、 y 的关系为 $z=0.2y-x$． 根据（II）的结果回答下列问题：

(i) 年宣传费 $x=49$ 时，年销售量及年利润的预报值是多少？
(ii) 年宣传费 x 为何值时，年利率的预报值最大？

附：对于一组数据 (u_1,v_1)，(u_2,v_2)，……，(u_n,v_n)，其回归线 $v=\alpha+\beta u$ 的斜率和截距的最小二乘估计分别为：

$$\beta=\frac{\sum_{i=1}^{n}(u_i-\overline{u})(v_i-\overline{v})}{\sum_{i=1}^{n}(u_i-\overline{u})^2},\alpha=\overline{v}-\beta\overline{u}.$$、

【答案】（I）$y=c+d\sqrt{x}$ 适合作为年销售 y 关于年宣传费用 x 的回归方程类型；（II）$\hat{y}=100.6+68\sqrt{x}$ (III) 46．24

分析：（I）由散点图及所给函数图像即可选出适合作为拟合的函数；（II）令 $w=\sqrt{x}$，先求出建立 y 关于 w 的线性回归方程，即可 y 关于 x 的回归方程；(III)（i）利用 y 关于 x 的回归方程先求出年销售量 y 的预报值，再根据年利率 z 与 x、y 的关系为 z=0．2y−x 即可年利润 z 的预报值；(ii) 根据（II）的结果知，年利润 z 的预报值，列出关于 x 的方程，利用二次函数求最值的方法即可求出年利润取最大值时的年宣传费用．

解析：

（I）由散点图可以判断， $y=c+d\sqrt{x}$ 适合作为年销售 y 关于年宣传费用 x 的回归方程类型．

（II）令 $w=\sqrt{x}$，先建立 y 关于 w 的线性回归方程，由于 $\hat{d}=\frac{\sum_{i=1}^{8}(w_i-\overline{w})(y_i-\overline{y})}{\sum_{i=1}^{8}(w_i-\overline{w})^2}=\frac{108.8}{16}=68$，

∴ $\hat{c}=\overline{y}-\hat{d}\overline{w}$=563−68×6．8=100．6．

∴ y 关于 w 的线性回归方程为 $\hat{y}=100.6+68w$，

∴ y 关于 x 的回归方程为 $\hat{y}=100.6+68\sqrt{x}$．

(III)（i）由（II）知，当 x=49 时，年销售量 y 的预报值

$\hat{y}=100.6+68\sqrt{49}$=576．6，

$\hat{z}=576.6\times 0.2-49=66.32$．

（ii）根据（II）的结果知，年利润 z 的预报值

$\hat{z}=0.2(100.6+68\sqrt{x})-x=-x+13.6\sqrt{x}+20.12$，

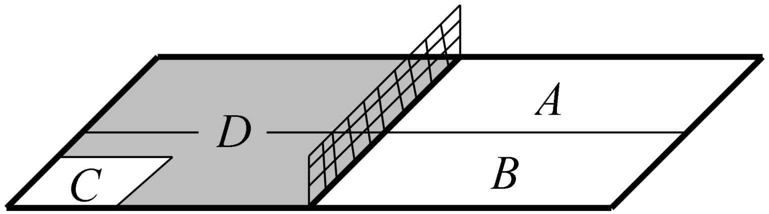
∴当 $\sqrt{x}=\dfrac{13.6}{2}=6.8$，即 $x=46.24$ 时，$\hat{z}$ 取得最大值．

故宣传费用为 46．24 千元时，年利润的预报值最大．……12 分

5.（2014 高考数学山东理科·第 18 题）乒乓球台面被球网分隔成甲、乙两部分，如图，甲上有两个不相交的区域 A,B，乙被划分为两个不相交的区域 C,D．某次测试要求队员接到落点在甲上的来球后向乙回球．规定：回球一次，落点在 C 上记 3 分，在 D 上记 1 分，其它情况记 0 分．对落点在 A 上的来球，队员小明回球的落点在 C 上的概率为 $\dfrac{1}{2}$，在 D 上的概率为 $\dfrac{1}{3}$；对落点在 B 上的来球，小明回球的落点在 C 上的概率为 $\dfrac{1}{5}$，在 D 上的概率为 $\dfrac{3}{5}$．假设共有两次来球且落在 A,B 上各一次，小明的两次回球互不影响．求：

（I）小明两次回球的落点中恰有一次的落点在乙上的概率；

（II）两次回球结束后，小明得分之和 ξ 的分布列与数学期望．



【答案】（I）$\dfrac{3}{10}$；（II）见解析

解析：（Ⅰ）记 A_i 为事件“小明对落点在 A 上的来球回球的得分为 i 分”（$i=0,1,3$） 则 $P(A_3)=\frac{1}{2},P(A_1)=\frac{1}{3},P(A_0)=1-\frac{1}{2}-\frac{1}{3}=\frac{1}{6}$；

记 B_i 为事件“小明对落点在 B 上的来球回球的得分为 i 分”（$i=0,1,3$） 则 $P(B_3)=\frac{1}{5},P(B_1)=\frac{3}{5},P(B_0)=1-\frac{1}{5}-\frac{3}{5}=\frac{1}{5}$；

记 D 为事件“小明两次回球的落点中恰有一次的落点在乙上”，

由题意，$D=A_3B_0+A_1B_0+A_0B_1+A_0B_3$，由事件的独立性及互斥性得

$$P(D)=P(A_3B_0+A_1B_0+A_0B_1+A_0B_3)=P(A_3B_0)+P(A_1B_0)+P(A_0B_1)+P(A_0B_3)$$

$$=P(A_3)P(B_0)+P(A_1)P(B_0)+P(A_0)P(B_1)+P(A_0)P(B_3)=\frac{1}{2}\times\frac{1}{5}+\frac{1}{3}\times\frac{1}{5}+\frac{1}{6}\times\frac{3}{5}+\frac{1}{6}\times\frac{1}{5}=\frac{3}{10}.$$

所以小明两次回球的落点中恰有一次的落点在乙上的概率为 $\frac{3}{10}$.

（Ⅱ）由题意，随机变量 ξ 可能的取值为 $0,1,2,3,4,6$．由事件的独立性及互斥性得

$P(\xi=0)=P(A_0B_0)=\frac{1}{6}\times\frac{1}{5}=\frac{1}{30}$；　$P(\xi=1)=P(A_1B_0+A_0B_1)=P(A_1B_0)+P(A_0B_1)=\frac{1}{3}\times\frac{1}{5}+\frac{1}{6}\times\frac{3}{5}=\frac{1}{6}$；

$P(\xi=2)=P(A_1B_1)=\frac{1}{3}\times\frac{3}{5}=\frac{1}{5}$；　$P(\xi=3)=P(A_3B_0+A_0B_3)=P(A_3B_0)+P(A_0B_3)=\frac{1}{2}\times\frac{1}{5}+\frac{1}{6}\times\frac{1}{5}=\frac{2}{15}$；

$P(\xi=4)=P(A_3B_1+A_1B_3)=P(A_3B_1)+P(A_1B_3)=\frac{1}{2}\times\frac{3}{5}+\frac{1}{3}\times\frac{1}{5}=\frac{11}{30}$；　$P(\xi=6)=P(A_3B_3)=\frac{1}{2}\times\frac{1}{5}=\frac{1}{10}$.

可得随机变量 ξ 的分布列为

ξ	0	1	2	3	4	6
P	$\frac{1}{30}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{2}{15}$	$\frac{11}{30}$	$\frac{1}{10}$

所以，数学期望 $E\xi=0\times\frac{1}{30}+1\times\frac{1}{6}+2\times\frac{1}{5}+3\times\frac{2}{15}+4\times\frac{11}{30}+6\times\frac{1}{10}=\frac{91}{30}$.

6.(2014 高考数学江西理科·第 22 题)随机将 $1,2,\cdots,2n\left(n\in N^*,n\geq 2\right)$ 这 2n 个连续正整数分成 A, B 两组, 每组 n 个数, A 组最小数为 a_1, 最大数为 a_2 ;B 组

最小数为 b_1，最大数为 b_2，记 $\xi=a_2-a_1,\eta=b_2-b_1$

(1) 当 $n=3$ 时，求 ξ 的分布列和数学期望；

(2) 令 C 表示事件 ξ 与 η 的取值恰好相等，求事件 C 发生的概率 $p(c)$；

(3) 对 (2) 中的事件 C，$\overline{c}$ 表示 C 的对立事件，判断 $p(c)$ 和 $p(\overline{c})$ 的大小关系，并说明理由.

【答案】(1)

ξ	2	3	4	5
P	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{1}{5}$

$E\xi=\frac{7}{2}$. (2) 当 $n=2$ 时，$P(C)=\frac{4}{6}=\frac{2}{3}$，当 $n\geq 3$ 时 $P(C)=\frac{2(2+\sum\limits_{k=1}^{n-2}C_{2k}^{k})}{C_{2n}^{n}}$

(3) 当 $n=2$ 时，$P(C)>P(\overline{C})$，当 $n\geq 3$ 时，$P(C)<P(\overline{C})$，

分析：(1) 当 $n=3$ 时，将 6 个正整数平均分成 A，B 两组，不同的分组方法共有 $C_6^3=20$ 种，ξ 所有可能值为 2，3，4，5．对应组数分别为 4，6，6，4，对应概率为 $\frac{1}{5}$，$\frac{3}{10}$，$\frac{3}{10}$，$\frac{1}{5}$，$E\xi=2\times\frac{1}{5}+3\times\frac{3}{10}+4\times\frac{3}{10}+5\times\frac{1}{5}=\frac{7}{2}$. (2) ξ 和 η 恰好相等的所有可能值为 $n-1,n,n+1,\cdots,2n-2$. 当 ξ 和 η 恰好相等且等于 $n-1$ 时，不同的分组方法有 2 种；当 ξ 和 η 恰好相等且等于 n 时，不同的分组方法有 2 种；当 ξ 和 η 恰好相等且等于 $n+1$ 时，不同的分组方法有 2 C_2^1 种；当 ξ 和 η 恰好相等且等于 $n+2$ 时，不同的分组方法有 2 C_4^2 种；以此类推：ξ 和 η 恰好相等且等于 $n+k(k=1,2,\cdots,n-2),(n\geq 3)$ 时，不同的分组方法有 2 C_{2k}^{k} 种；所以当 $n=2$ 时，$P(C)=\frac{4}{6}=\frac{2}{3}$

当 $n\geq 3$ 时 $P(C)=\dfrac{2(2+\sum\limits_{k=1}^{n-2}C_{2k}^{k})}{C_{2n}^{n}}$ (3) 先归纳：当 $n=2$ 时，$P(\overline{C})=\dfrac{1}{3}$, 因此 $P(C)>P(\overline{C})$, 当 $n\geq 3$ 时，$P(C)<P(\overline{C})$, 即证当 $n\geq 3$ 时 $P(C)=\dfrac{2(2+\sum\limits_{k=1}^{n-2}C_{2k}^{k})}{C_{2n}^{n}}<\dfrac{1}{2}$, $4(2+\sum\limits_{k=1}^{n-2}C_{2k}^{k})<C_{2n}^{n}$，这可用数学归纳法证明．当 $n=m+1$ 时，$4(2+\sum\limits_{k=1}^{m+1-2}C_{2k}^{k})=4(2+\sum\limits_{k=1}^{m-2}C_{2k}^{k})+4C_{2m+2}^{m+1}<C_{2m}^{m}+4C_{2m+2}^{m+1}$，利用阶乘作差 $C_{2m+2}^{m+1}-(C_{2m}^{m}+4C_{2m+2}^{m+1})$ 可得大小.

解析：(1) 当 $n=3$ 时，ξ 所有可能值为 2, 3, 4, 5．将 6 个正整数平均分成 A, B 两组，不同的分组方法共有 $C_6^3=20$ 种，所以 ξ 的分布列为

ξ	2	3	4	5
P	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{1}{5}$

$$E\xi=2\times\frac{1}{5}+3\times\frac{3}{10}+4\times\frac{3}{10}+5\times\frac{1}{5}=\frac{7}{2}.$$

(2) ξ 和 η 恰好相等的所有可能值为 $n-1,n,n+1,\cdots,2n-2$.

又 ξ 和 η 恰好相等且等于 $n-1$ 时，不同的分组方法有 2 种；ξ 和 η 恰好相等且等于 n 时，不同的分组方法有 2 种；

ξ 和 η 恰好相等且等于 $n+k(k=1,2,\cdots,n-2),(n\geq 3)$ 时，不同的分组方法有 $2C_{2k}^{k}$ 种；所以当 $n=2$ 时，$P(C)=\dfrac{4}{6}=\dfrac{2}{3}$

当 $n\geq 3$ 时 $P(C)=\dfrac{2(2+\sum\limits_{k=1}^{n-2}C_{2k}^{k})}{C_{2n}^{n}}$

(3) 由 (2) 当 $n=2$ 时，$P(\overline{C})=\frac{1}{3}$，因此 $P(C)>P(\overline{C})$，而当 $n\geq 3$ 时，$P(C)<P(\overline{C})$，理由如下：

$P(C)<P(\overline{C})$，等价于 $4(2+\sum_{k=1}^{n-2}C_{2k}^{k})<C_{2n}^{n}$ ①

用数学归纳法来证明：

1° 当 $n=3$ 时，①式左边 $=4(2+C_2^1)=16$，①式右边 $=C_6^3=20$，所以①式成立

2° 假设 $n=m(m\geq 3)$ 时①式成立，即 $4(2+\sum_{k=1}^{m-2}C_{2k}^{k})<C_{2m}^{m}$ 成立

那么，当 $n=m+1$ 时，①式左边 $=4(2+\sum_{k=1}^{m+1-2}C_{2k}^{k})=4(2+\sum_{k=1}^{m-2}C_{2k}^{k})+4C_{2m+2}^{m+1}<C_{2m}^{m}+4C_{2m+2}^{m+1}$

$$=\frac{(2m)!}{m!m!}+\frac{4\times(2m-2)!}{(m-1)!(m-1)!}=\frac{(m+1)^2(2m)(2m-2)!(4m-1)}{(m+1)!(m+1)!}$$

$$<\frac{(m+1)^2(2m)(2m-2)!(4m)}{(m+1)!(m+1)!}=C_{2m+2}^{m+1}\cdot\frac{2(m+1)m}{(2m+1)(2m-1)}<C_{2m+2}^{m+1}$$ =①式右边

即当 $n=m+1$ 时①式也成立

综合 1° 2° 得，对于 $n\geq 3$ 的所有正整数，都有 $P(C)<P(\overline{C})$ 成立．