概率论与数理统计第一次习题课题目解答

题1 从一批产品中任取n件，以事件A_i表示“第i件取得正品”，用它们表示下列事件：

1. 没有一件是次品(全是正品)
   答：$\bigcap\limits_{i=1}^{n} A_i$。
   如果用示性函数表达，则该事件为
   $$I_{A_1}+\cdots+I_{A_n}=n$$
   或等价地，
   $$I_{A_1}\cdots I_{A_n}=1.$$

2. 至少有一件是次品
   答：直接的表示为$\bigcup\limits_{i=1}^{n} \overline{A_i}$，或用对偶律，表示为$\overline{\bigcap\limits_{i=1}^{n} A_i}$。
   如果用示性函数表达，则该事件为
   $$I_{A_1}+\cdots+I_{A_n}<n.$$

3. 仅仅只有一件是次品
   答：$\bigcup\limits_{i=1}^{n}\left(\overline{A_i}\cdot\bigcap\limits_{1\le j\le n, j\ne i} A_j\right)$，或者$\overline{\left(A_1\cdots A_n\right)\cup\bigcup\limits_{1\le i<j\le n}\overline{A_i}\cdot\overline{A_j}}$，这两个形式不同的表达你喜欢哪个？如果用示性函数表达，则该事件为
   $$I_{A_1}+\cdots+I_{A_n}=n-1.$$

4. 至少有两件不是次品
   答：直接表达为$\bigcup\limits_{1\le i,j\le n; i\ne j}\left(A_iA_j\right)$，也可以间接表达为
   $$\overline{\left(\overline{A_1}\cdots\overline{A_n}\right)\cup\bigcup_{1\le i\le n}\left(A_i\bigcap_{1\le j\le n; j\ne i}\overline{A_j}\right)}.$$

   第一个表达形式简单而且直接，但求和的事件们不是互不相容的，计算概率时会较麻烦；第二个表达虽形式复杂，但表示对立的横线下已表达为互不相容的一些事件，便于

概率计算。表达事件的目的是为随后的概率计算提供方便，因此样子略显古怪的后者比样子简单的前者更好。如果用示性函数表达，则该事件为

$$I_{A_1}+\cdots+I_{A_n}\geq 2.$$

题2 设有来自2个地区的考生的报名表，其中第k个地区男女生报名表分别有b_k份和g_k份，$k=1,2$。随机地取一个地区的报名表，从中先后抽取两份,求：

1. 先抽到的一份是女生表的概率。

   解：记D_i为事件“抽到第i个地区”，G_j为事件“第j次抽到女生”，B_j为事件“第j次抽到男生”（首先交待要使用的符号的含义）。
   按考生的来源进行分类，用全概率公式

$$\begin{aligned}P(G_1)&=P(D_1)P(G_1|D_1)+P(D_2)P(G_1|D_2)\quad\text{（先用事件符号说明数量之间的关系）}\\&=\frac{1}{2}\times\frac{g_1}{b_1+g_1}+\frac{1}{2}\times\frac{g_2}{b_2+g_2}=\frac{g_1(b_2+g_2)+g_2(b_1+g_1)}{2(b_1+g_1)(b_2+g_2)}.\end{aligned}$$

2. 已知后抽到的一份是男生表，求先抽到的一份是女生表的概率

解法1：

$$\begin{aligned}P(G_1|B_2)&=\frac{P(G_1B_2)}{P(B_2)}\\&=\frac{\sum\limits_{i=1}^{2}P(D_iG_1B_2)}{P(B_2)}\\&=\frac{\sum\limits_{i=1}^{2}P(D_i)P(G_1|D_i)P(B_2|D_iG_1)}{P(B_1)}\\&=\frac{\sum\limits_{i=1}^{2}\frac{1}{2}\times\frac{g_i}{b_i+g_i}\times\frac{b_i}{b_i+g_i-1}}{\sum\limits_{i=1}^{2}\frac{1}{2}\times\frac{b_i}{b_i+g_i}}\\&=\frac{b_1g_1(b_2+g_2)(b_2+g_2-1)+b_2g_2(b_1+g_1)(b_1+g_1-1)}{(b_1+g_1-1)(b_2+g_2-1)[b_1(b_2+g_2)+b_2(b_1+g_1)]},\end{aligned}$$

其中$P(B_2)$利用了抽签模型与次序无关的性质，所以$P(B_2) = P(B_1)$。大家也可以用全概率公式计算$P(G_2)$，会得到同样的结果。上面第三个等号右端分子部分的计算中，我们使用了双重条件的全概率公式。

解法2： 有的同学会用全概率公式这样做：

$$P(G_1|B_2) = \frac{1}{2}\frac{g_1}{b_1+g_1-1} + \frac{1}{2}\frac{g_2}{b_2+g_2-1},$$

并解释说，1/2是挑选一个地区的概率，而在第k个地区里，如果已知后选中一个男孩，那么第一个人就只能来自于剩下的b_k+g_k-1个人，而其中女生有g_k个。
这个解释直观上好像是对的，但它是错的，因为已知后一个是男生的条件下，两个地区被选中的机会就不再一样了，一个极端的例子是第一个地区根本就没有男生，这时这个错误是明显的。
正确的方式是这样的：

$$\begin{aligned}P(G_1|B_2) &= \sum_{i=1}^{2} P(D_i|B_2)P(G_1|D_iB_2) \quad (\text{这是一个用条件概率表达的全概率公式}) \\ &= P(D_1|B_2)\frac{g_1}{b_1+g_1-1} + P(D_2|B_2)\frac{g_2}{b_2+g_2-1}, \\ &\qquad (\text{注意到}P(G_1|D_iB_2) = P(G_2|D_iB_1))\end{aligned}$$

其中

$$P(D_1|B_2) = \frac{P(D_1)P(B_2|D_1)}{\sum\limits_{k=1}^{2} P(D_k)P(B_2|D_k)} = \frac{\frac{1}{2}\times\frac{b_1}{b_1+g_1}}{\sum\limits_{k=1}^{2}\frac{1}{2}\times\frac{b_k}{b_k+g_k}} = \frac{b_1(b_2+g_2)}{b_1(b_2+g_2)+b_2(b_1+g_1)},$$

同理可得$P(D_2|B_2)$的值，代入第一个等式得到与解法1相同的结论。
开始提到的那个错误就是错误地认为$P(D_i|B_2) = P(D_i)$，而忽视了已经发生的事件对选择地区时概率的影响。注意到在上述全概率公式中，B_2是等号左端要求的条件概率中的条件，它应该是整个问题讨论的共同前提，因此要作为条件出现在等号右端出现的每一个条件概率中。

3. 假设不先确定一个地区，而是从所有报名表中随机抽取两份。如果已知后抽到的一份是一个男生的报名表，那么问先抽到的一份是同地区一个女生的报名表的可能性有多大？
解：由于不预选地区，所以

$$P(B_2) = P(B_1) = \frac{b_1+b_2}{b_1+g_1+b_2+g_2}.$$

记C_i为事件“两份表格都来自第i个地区”。则事件“两份表格来自同一地区”为$C_1 \cup C_2$，

$$
\begin{aligned}
P((C_1 \cup C_2)G_1|B_2) &= \frac{P((C_1 \cup C_2)G_1B_2)}{P(B_2)} \\
&= \frac{\sum_{i=1}^{2} P(C_iG_1B_2)}{P(B_2)} \\
&= \frac{\frac{g_1b_1}{(b_1+g_1+b_2+g_2)(b_1+g_1+b_2+g_2-1)} + \frac{g_2b_2}{(b_1+g_1+b_2+g_2)(b_1+g_1+b_2+g_2-1)}}{\frac{b_1+b_2}{b_1+g_1+b_2+g_2}} \\
&= \frac{g_1b_1 + g_2b_2}{(b_1+g_1+b_2+g_2-1)(b_1+b_2)}.
\end{aligned}
$$

题3 口袋中有a个黑球和b个白球，每次从口袋中随机地摸出一球，并换成一个黑球。

1. 问第k次摸球时，摸到黑球的概率是多少？

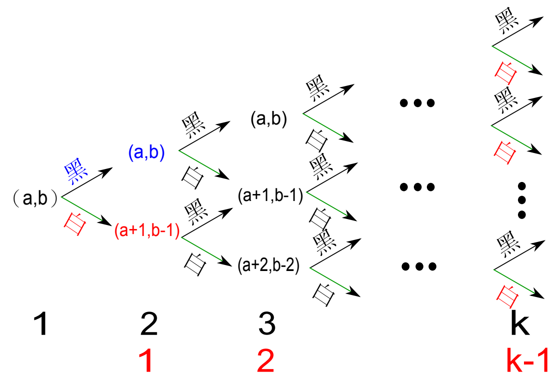
解法1： 记$A_k(a,b)$表示“从最初有a个黑球和b个白球的袋中按上述规则取球，第k次取到黑球”，用首步分析法：

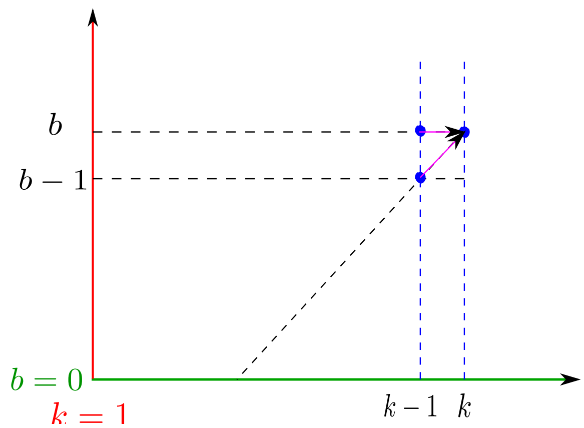
$$
\begin{aligned}
&P(\overline{A_k(a,b)}) \\
=&P(A_1(a,b))P(\overline{A_k(a,b)}|A_1(a,b)) + P(\overline{A_1(a,b)})P(\overline{A_k(a,b)}|\overline{A_1(a,b)}) \\
=&\frac{a}{a+b}P(\overline{A_{k-1}(a,b)}) + \frac{b}{a+b}P(\overline{A_{k-1}(a+1,b-1)}) \qquad (\text{为什么？})
\end{aligned}
$$

因为总球数$N=a+b$固定不变，我们可以记$p(k,b)=P(\overline{A_k(a,b)})$。则上述递推写成

$$
p(k,b) = \left(1 - \frac{b}{N}\right)p(k-1,b) + \frac{b}{N}p(k-1,b-1).
$$

这是一个双重指标的迭代，在(k,b)坐标平面上不难看出上述迭代在指标之间的依赖关系。





利用边界条件$p(k,0)=0$和初始条件$p(1,b)=\frac{b}{a+b}$借助上述递推关系可以算得

$$p(2,b)=\frac{b}{a+b}\cdot\frac{a+b-1}{a+b},\qquad p(3,b)=\frac{b}{a+b}\cdot\left(\frac{a+b-1}{a+b}\right)^2.$$

从而猜想$p(k,b)=\frac{b}{a+b}\cdot(\frac{a+b-1}{a+b})^{k-1}$，然后用数学归纳法证明猜想成立。所以，

$$P(A_k(a,b))=1-p(k,b)=1-\frac{b}{a+b}\cdot\left(1-\frac{1}{a+b}\right)^{k-1}.$$

解法2 以白球做为切入点（为什么呢？），把b个白球分别编号为$1,2,\cdots,b$，定义

$B_{k,i}$：第k次摸球时摸到的是编号为i的白球（$i=1,2,\cdots,b$）

我们还是先从事件关系入手：

$$\overline{A_k}=B_{k,1}\cup B_{k,2}\cup B_{k,3}\cup\cdots\cup B_{k,b},\qquad \text{摸到白球，即摸到某个白球}$$

$$B_{k,i}B_{k,j}=\emptyset,\quad \forall i\neq j,\qquad \text{一次不会摸到两个白球}$$

$$B_{k,i}=\overline{B_{1,i}}\cdot\overline{B_{2,i}}\cdot\overline{B_{3,i}}\cdots\overline{B_{k-1,i}}\cdot B_{k,i}\qquad \text{第}k\text{次摸到的白球，此前一定从未被摸到.}$$

于是，

$$
\begin{aligned}
P(\overline{A_k}) &= \sum_{i=1}^{b} P(B_{k,i}) \\
P(B_{k,i}) &= P\left(\overline{B_{1,i}} \cdot \overline{B_{2,i}} \cdot \overline{B_{3,i}} \cdots \overline{B_{k-1,i}} \cdot B_{k,i}\right) \\
&= P\left(\overline{B_{1,i}}\right) P\left(\overline{B_{2,i}} | \overline{B_{1,i}}\right) P(\overline{B_{3,i}} | \overline{B_{1,i}} \cdot \overline{B_{2,i}}) \cdots P\left(B_{k,i} | \overline{B_{1,i}} \cdot \overline{B_{2,i}} \cdot \overline{B_{3,i}} \cdots \overline{B_{k-1,i}}\right) \\
&= \left(\frac{a+b-1}{a+b}\right)^{k-1} \frac{1}{a+b} \\
P(\overline{A_k}) &= \left(\frac{a+b-1}{a+b}\right)^{k-1} \frac{b}{a+b} \\
P(A_k) &= 1 - P(\overline{A_k}) = 1 - \left(\frac{a+b-1}{a+b}\right)^{k-1} \frac{b}{a+b}
\end{aligned}
$$

2. 以下给出第1问的一个解答，请判断这种方法对不对。
用“末步分析法”（即用此前过程的最后一步的不同结果作划分，使用全概率公式）

$$P(A_k) = P(A_{k-1})P(A_k|A_{k-1}) + P(\overline{A_{k-1}})P(A_k|\overline{A_{k-1}})$$

如果第$k-1$次取得黑球，那么第k次取球前袋中的情况与第$k-1$次取球前袋中情况完全相同，于是

$$P(A_k|A_{k-1}) = P(A_{k-1});$$

如果第$k-1$次取得白球，那么新换入的那个黑球导致第k次取到黑球的机会比上一次取得黑球的机会增加了$1/(a+b)$，于是

$$P(A_k|\overline{A_{k-1}}) = P(A_{k-1}) + \frac{1}{a+b}.$$

因此

$$P(A_k) = P(A_{k-1})^2 + P(\overline{A_{k-1}})\left(P(A_{k-1}) + \frac{1}{a+b}\right).$$

然后再此基础上利用初值和数学归纳法可以得到第1问中的结论。
在这个解法中，关于两个条件概率值的解释貌似很直观，但却是错的，我们被错误的直

观所蒙蔽了。事实上，

$$P(A_2) = P(A_1A_2) + P(A_1^cA_2) = \frac{a}{a+b} \times \frac{a}{a+b} + \frac{b}{a+b} \times \frac{a+1}{a+b} = \frac{a^2 + b(a+1)}{(a+b)^2},$$

$$\begin{aligned} P(A_2A_3) = P(A_1A_2A_3) + P(A_1^cA_2A_3) &= \frac{a}{a+b} \times \frac{a}{a+b} \times \frac{a}{a+b} + \frac{b}{a+b} \times \frac{a+1}{a+b} \times \frac{a+1}{a+b} \\ &= \frac{a^3 + b(a+1)^2}{(a+b)^3}, \end{aligned}$$

由此，我们可以得到

$$P(A_3|A_2) - P(A_2) = \frac{ab}{[a^2 + b(a+1)]\,(a+b)^2 \geq 0},$$

并且$P(A_3|A_2) \geq P(A_2)$，而且等号成立当且仅当$a = 0$或$b = 0$。
当$b = 0$时，自始至终没有白球，这时A_k是必然事件，因此

$$P(A_k|A_{k-1}) = P(A_k) = 1 = P(A_{k-1}), \quad \forall k > 1.$$

当$a = 0, b = 1$时，即开始时袋中只有一个球，而且是白球，这时A_1是不可能事件，$A_k(k > 1)$是必然事件，所以

$$P(A_k|A_{k-1}) = P(A_k) = 1 = P(A_{k-1}), \quad \forall k > 2.$$

当$a = 0, b > 1$时，即开始时袋中只有白球且多于1个，这时A_1是不可能事件，这时

$$P(A_3|A_2) = P(A_2), \quad P(A_4|A_3) > P(A_3).$$

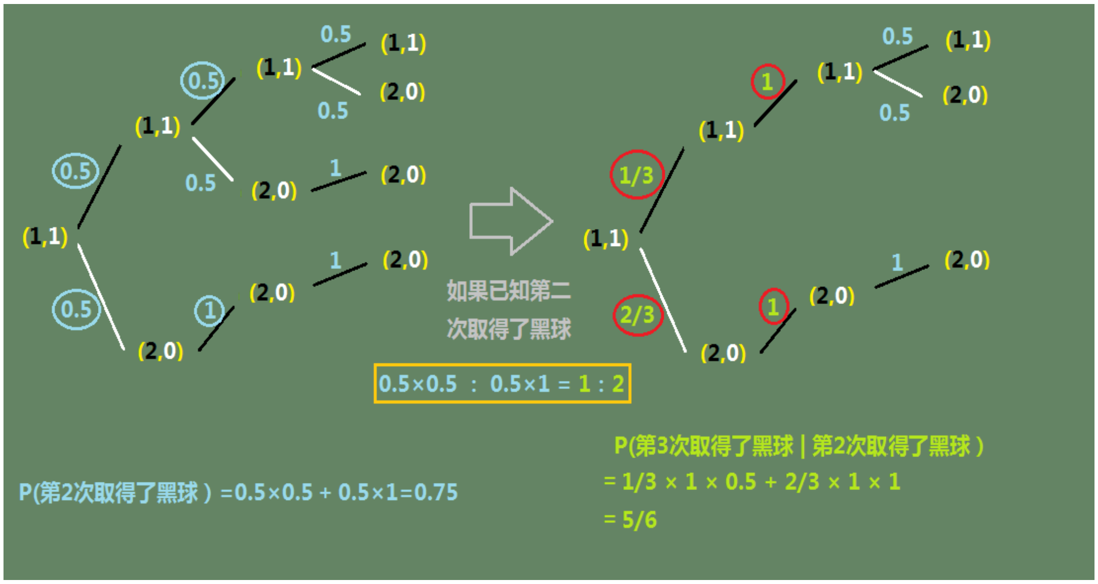
我们证明：如果最初袋中至少有2个球，且不全是黑球，那么

- $P(A_k|A_{k-1}) \geq P(A_{k-1})$，$\forall k > 2$；
- 上述不等式中等号成立，当且仅当$k = 3$且袋中最初全是白球；
- 如果袋中最初有黑球，则$P(A_2|A_1) = P(A_1)$。

最后一条很容易经计算得到验证。我们只看前两条，我们无妨假定最初袋中一共有N个球，且既有黑球也有白球，我们来证明

$$P(A_k|A_{k-1}) > P(A_{k-1}), \quad k \geq 3.$$

这时从最初的状态出发，第一次我们可能取得黑球，也可能取得白球，整个过程被分为两枝，并且最终成为一棵树，这棵树的每一个节点对应某次取球前袋中的状态，这次取球对应一条边，我们根据取出的球的颜色把这条边染成相应的颜色，于是这棵树的每条边被染成黑白两种颜色之一。



假设B是从根节点（对应袋子最初的状态）出发一条长度为$k-2$的路径，它对应前$k-2$次一个特定次序的取球过程，记$y(B)$表示路径B的终点所对应的状态中黑球的个数。记$\mathcal{B}_{k-2}$是从根节点出发长度为$k-2$的所有路径的全体。于是

$$P(A_{k-1}) = \sum_{B\in\mathcal{B}_{k-2}} P(B)P(A_{k-1}|B) = \sum_{B\in\mathcal{B}_{k-2}} P(B)\times\frac{y(B)}{N},$$

$$P(A_{k-1}A_k) = \sum_{B\in\mathcal{B}_{k-2}} P(B)P(A_{k-1}|B)P(A_k|BA_{k-1}) = \sum_{B\in\mathcal{B}_{k-2}} P(B)\times\frac{y(B)}{N}\times\frac{y(B)}{N}.$$

由Cauchy-Schwarz不等式知，

$$\begin{aligned}
[P(A_{k-1})]^2 &= \left(\sum_{B\in\mathcal{B}_{k-2}} \sqrt{P(B)}\times\sqrt{P(B)}\times\frac{y(B)}{N}\right)^2\\
&\le \left(\sum_{B\in\mathcal{B}_{k-2}} P(B)\right)\times\left(\sum_{B\in\mathcal{B}_{k-2}} P(B)\times\left(\frac{y(B)}{N}\right)^2\right)\\
&= P(A_{k-1}A_k),
\end{aligned}$$

等号成立当且仅当$\frac{y(B)}{N}$是不依赖于$B\in\mathcal{B}_{k-2}$的常数，但这是不可能的，所以

$$[P(A_{k-1})]^2 < P(A_{k-1}A_k),$$

故$P(A_k|A_{k-1}) > P(A_{k-1})$。也就是说，在整个取球过程中，一旦我们取出黑球，虽然我们把它放回使袋中球的状态恢复到这次取球前的样子，但是这时我们再取，得到黑球的机会就增加了。这和你的直观想象是不是很不一样？但它却是事实。因此，在讨论问题时，我们可以利用直观帮助我们思考，但是仅依靠所谓直观有可能犯错误，这就是一个例证。这时因为，在概率问题中除了“状态”还有相应的“概率”，一样的状态可能在不同情形下对应不同的概率，后者不象前者那样显而易见。因此，在学习概率论时要更加谨慎思考。

评论 首步分析法和末步分析法本身在逻辑上都没有问题，但针对具体问题，不是二者都适用，要视具体问题的特殊性，套用一句套话“要把全概率公式的普遍真理与具体问题的实际情况相结合”。

题5 口袋中有N个球，分别标有号码$1 \sim N$，现从中任取m个$(m < N)$，记最小值为X，最大值为Y。

1. 取球不返回时，写出X、Y的分布列。

   解：

$$
\begin{aligned}
P(X=k) &= P(X>k-1) - P(X>k) \quad \text{（对最小值这伎俩常有效）} \\
&= \frac{C_{N-(k-1)}^m}{C_N^m} - \frac{C_{N-k}^m}{C_N^m} = \frac{C_{N-k}^{m-1}}{C_N^m}, \quad k=1,\cdots,N-m+1.
\end{aligned}
$$

   其中，最后一个等号利用了杨辉（中国宋代智者，不知比西方早了多少个甲子）发现的等式$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$。

$$
\begin{aligned}
P(Y=k) &= P(Y<k+1) - P(Y<k) \quad \text{（对最大值这伎俩很管用）} \\
&= \frac{C_k^m}{C_N^m} - \frac{C_{k-1}^m}{C_N^m} = \frac{C_{k-1}^{m-1}}{C_N^m}, \quad k=m,\cdots,N.
\end{aligned}
$$

2. 取球返回时，写出X、Y的分布列。

   解：现在不过是独立重复试验罢了，招数与上面一样。

$$
\begin{aligned}
P(X=k) &= P(X>k-1) - P(X>k) = \frac{(N-k+1)^m}{N^m} - \frac{(N-k)^m}{N^m}, \\
P(Y=k) &= P(Y<k+1) - P(Y<k) = \frac{k^m}{N^m} - \frac{(k-1)^m}{N^m}.
\end{aligned}
$$

   两个随机变量的取值范围都是$k=1,\cdots,N$。

注：对一个离散概率分布列$\{(x_k,p_k)\}_k$，我们验证它正确性的一个方法（必要条件）是$\sum_k p_k = 1$。请对以上分布列验证这个条件。

题6 抽查一个家庭，考察两个事件. A: 至多有一个女孩; B: 男女都有. 针对下面两类家庭, 讨论事件是否独立:

1. 3 个孩子之家;

   解： 在 3 个孩子之家，以长幼顺序写出孩子性别，则由 (男、男、男), (男、男、女), (男、女、男), (女、男、男), (男、女、女), (女、男、女), (女、女、男), (女、女、女), 共 8 种可能情况作为样本空间。若假定男女出生率一样，则各样本点出现的概率均为 $\frac{1}{8}$ A 的有利场合为前 4 个样本点, B 的有利场合为当中的 6 个样本点，故 $P(A) = \frac{4}{8}$, $P(B) = \frac{6}{8}$. 而 AB 有利场合为第 2, 第 3, 第 4 个样本点，故 $P(AB) = \frac{3}{8}$. 这时有

   $$P(AB) = \frac{3}{8} = \frac{4}{8} \times \frac{6}{8} = P(A) \times P(B).$$

   因此 A 与 B 独立。

2. 4 个孩子之家.

   解： 在 4 个孩子之家，计有 $2^4 = 16$ 个样本点，也等可能。不难算得

   $$P(A) = \frac{5}{16} \text{ (男孩1 种，或3 男1 女，而1 女可以是任何排行，共4 种，因此共有5 种)}$$
   $$P(B) = \frac{14}{16} \text{ (扣去全男或全女之后的14 种)}$$
   $$P(AB) = \frac{4}{16} \text{ (3 男1 女共4 种)}$$

   这说明 A 与 B 不独立。

3. 如果是 n 个孩子呢？

   解： 在 n 个孩子之家，计有 2^n 个样本点，亦为等可能模型。容易算得

   $$P(A) = \frac{1}{2^n} + \frac{n}{2^n} = \frac{n+1}{2^n} \text{ (全为男孩和有一个女孩)}$$
   $$P(B) = 1 - \frac{1}{2^n} - \frac{1}{2^n} \text{ (1 减去全为男孩和全为女孩的的概率)}$$
   $$P(AB) = \frac{n}{2^n} \text{ (A, B相交就是正好有一个女孩的情况)}$$

于是有

$$P(AB) - P(A)P(B) = \frac{2^{n-1} - (n+1)}{2^{2n-1}}$$

稍微考察 $y = x + 1$ 与 $y = 2^{x-1}$ 两个函数可知， $P(AB) = P(A)P(B)$ 当且仅当 $n = 3$. 所以 A 与 B 仅在有3 个孩子的情况时独立，其余情况下不独立。

4. 当 $n \neq 3$ 时，事件 A, B 是相互促进还是相互抑制？利用其相关系数进行说明。

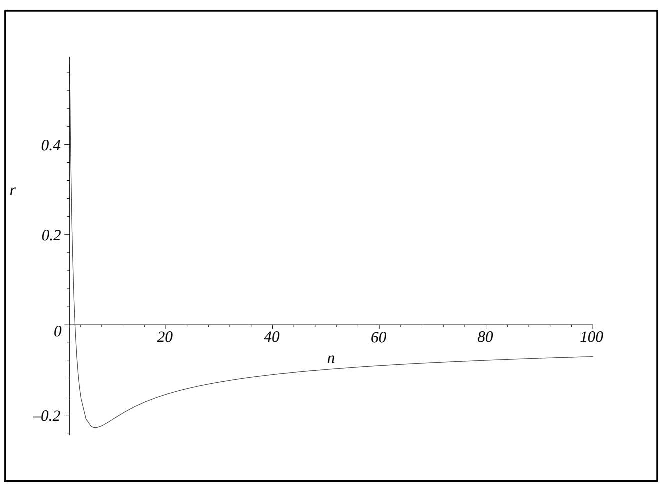
解： A, B 的相关系数为：

$$\begin{aligned} r_{A,B} &= \frac{P(AB) - P(A)P(B)}{\sqrt{P(A)P(A^c)P(B)P(B^C)}} = \frac{\frac{n}{2^n} - \frac{n+1}{2^n}(1 - \frac{1}{2^{n-1}})}{\sqrt{\frac{n+1}{2^n}(1 - \frac{n+1}{2^n})(1 - \frac{1}{2^{n-1}})\frac{1}{2^{n-1}}}} \\ &= \frac{n + 1 - 2^{n-1}}{2^{2n-1} \cdot \sqrt{\frac{n+1}{2^{2n-1}}(1 - \frac{n+1}{2^n})(1 - \frac{1}{2^{n-1}})}} \end{aligned}$$

当 $n < 3$ 时， $r_{A,B} > 0$, 此时 A, B 相互促进; 当 $n > 3$ 时， $r(A, B) < 0$, 此时 A, B 相互抑制。

上述例子说明，有时候直观并不完全可靠。

以下是相关系数随n的变化的图像



从这个图上我们会看到，$n > 3$时，A, B两个事件相互抑制，但随着n不断增大，这种抑制作用会越来越弱，当n足够大以后，二者又很接近于独立了。对这样的现象，你能解释吗？

题6 从装有m个白球、n个黑球的袋中不放回地取球，直到摸出白球时停止，记X为取出的黑球的个数，求X的分布。

解： 记事件B_k为“第k次摸到黑球”，事件W_k为“第k次摸到白球”，则

$$\begin{aligned}P(X=k) &= P(B_1\cdots B_k W_{k+1})\\ &= \frac{n}{m+n}\cdot\frac{n-1}{m+n-1}\cdots\frac{n-(k-1)}{m+n-(k-1)}\cdot\frac{m}{m+n-k}\\ &= \frac{n!}{(n-k)!}\frac{(m+n-k)!}{(m+n)!}\frac{m}{m+n-k}\\ &= \frac{C_n^k}{C_{m+n}^k}\frac{m}{m+n-k}\end{aligned}$$

题7 独立重复抛硬币，$A=$“抛一次得正面”，$p=P(A)$，记X为最早得到累计2个正面时抛硬币的次数，Y为最早得到连续2个正面时抛硬币的次数。求X,Y各自的概率分布。

解： 最早得到累计2个正面时抛硬币数为k说明前$k-1$次中有一次是正面，剩下$k-2$次是反面，第k次是正面，所以
$P(X=k)=C_{k-1}^1p^2(1-p)^{k-2}$

记$p_k=P(Y=k)$，记A_i为第i次摸到正面，对p_k应用首步分析法，当$k>2$时，

$$p_k=P(Y=k)=P(\overline{A_1})P(Y=k|\overline{A_1})+P(A_1A_2)P(Y=k|A_1A_2)+P(A_1\overline{A_2})P(Y=k|A_1\overline{A_2})$$

其中$P(Y=k|A_1A_2)=0$，因为$k>2$，第1、2次都摸到正面，不可能第k次最早摸到连续2个正面；$P(Y=k|\overline{A_1})=P(Y=k-1)$，因为第1次摸到反面，而且之后的游戏规则没有变；$P(Y=k|A_1\overline{A_2})=P(Y=k-2)$，因为第1次摸到正面，第2次摸到反面，之后的游戏规则也没有变。所以

$$p_k=p(1-p)p_{k-2}+(1-p)p_{k-1}$$

下面解这个差分方程：将$p_k=\lambda^k(\lambda\neq 0)$代入上式，得到特征方程

$$\lambda^2-(1-p)\lambda-p(1-p)=0,$$

解得

$$\lambda_{1,2}=\frac{(1-p)\pm\sqrt{(1-p)(1-5p)}}{2}.$$

如果$p\neq\frac{1}{5}$，则$\lambda_1\neq\lambda_2$，这时差分方程的通解为

$$p_k=C_1\lambda_1^k+C_2\lambda_2^k,$$

再由初始条件

$$p_2 = p^2, \qquad p_3 = 2(1-p)p^2,$$

确定线性组合系数C_1, C_2（过程和结果略）。

如果$p = \frac{1}{5}$，这时$\lambda_1 = \lambda_2 = \frac{2}{5}$则这时差分方程的通解为

$$p_k = (C_1 + C_2 k)\left(\frac{2}{5}\right)^k,$$

再由初始条件

$$p_2 = p^2, \qquad p_3 = 2(1-p)p^2,$$

确定线性组合系数C_1, C_2（过程和结果略）。

题8 （补充作业题）著名数学家Von Neumann说，即使用一枚不均匀的硬币也可以得到公平的机会。他的做法是：把这枚硬币掷两次，如果两次掷得的结果相同（都是正面或者都是反面），那么就接着再掷两次，直到出现两次结果不同，如果出现的是“正反”，我们定义这种情况为“赢”，如果出现的是“反正”，我们定义这种情况是“输”。

1. 证明P(抛硬币在有限次结束) $= 1$；
2. 证明P(最终赢) $= P$(最终输) $= 0.5$。

解： 首先设掷2次硬币为掷一轮，记事件A_n为“至第n轮以赢结束”，事件B_n为“至第n轮以输结束”，事件C_n为“到第n轮掷硬币还没有结束”，事件C为“抛硬币一直没有结束”，事件A为“最终赢”，事件B为“最终输”，则

$$\begin{aligned} P(C_n) &= [(p^2 + (1-p)^2)]^n, \\ P(A_n) &= [(p^2 + (1-p)^2)]^{n-1} p(1-p), \\ P(B_n) &= [(p^2 + (1-p)^2)]^{n-1} (1-p)p. \end{aligned}$$

注意到

$$C_1 \supset C_2 \supset \cdots \supset C_n \supset \cdots,$$

故

$$C = \bigcap_{n \geq 1} C_n = \lim_{n\to\infty} C_n,$$

因此

$$P(C) = \lim_{n\to\infty} P(C_n) = \lim_{n\to\infty} [(p^2 + (1-p)^2)]^n = 0,$$

即P(抛硬币在有限次结束) $= 1$。

而

$$P(A) = \sum_{n\geq 1} P(A_n) = \sum_{n\geq 1} P(B_n) = P(B),$$

又

$$P(A) + P(B) + P(C) = 1,$$

所以$P(A) = P(B) = 0.5$。这说明，Von Neumann的建议是可行的（几乎总会在有限轮分出输赢），并且输赢机会相等。