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：$\bigcap_{i=1}^{n} A_i$。
   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$\bigcup_{i=1}^{n}\overline{A_i}$，或用对偶律，表示为$\overline{\bigcap_{i=1}^{n} A_i}$。
   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：$\bigcup_{i=1}^{n}\left(\overline{A_i}\cdot\bigcap_{1\le j\le n, j\neq i} A_j\right)$，或者$\overline{(A_1\cdots A_n)\cup\bigcup_{1\le i<j\le n}\overline{A_i}\cdot\overline{A_j}}$，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$\bigcup_{1\le i,j\le n; i\neq j}(A_iA_j)$，也可以间接表达为
   $$\overline{(\overline{A_1}\cdots\overline{A_n})\cup\bigcup_{1\le i\le n}\left(A_i\bigcap_{1\le j\le n; j\neq i}\overline{A_j}\right)}.$$
   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
   $$I_{A_1}+\cdots+I_{A_n}\ge 2.$$

题2 设有来自3个地区各10名、15名和25名考生的报名表，其中女生的报名表分别为3份、7份和5份，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

$$
\begin{aligned}
P(G_1) &= P(D_1)P(G_1|D_1) + P(D_2)P(G_1|D_2) + P(D_3)P(G_1|D_3) \\
&\qquad \text{（先用事件符号说明数量之间的关系）} \\
&= \frac{1}{3}\times\frac{3}{10} + \frac{1}{3}\times\frac{7}{15} + \frac{1}{3}\times\frac{5}{25} \\
&\qquad \text{（再代入具体数值进行计算）} \\
&= \frac{29}{90}.
\end{aligned}
$$

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：

$$
\begin{aligned}
P(G_1|B_2) &= \frac{P(G_1B_2)}{P(B_2)} \\
&= \frac{\sum_{i=1}^{3} P(D_iG_1B_2)}{P(B_2)} \\
&= \frac{\sum_{i=1}^{3} P(D_i)P(G_1|D_i)P(B_2|D_iG_1)}{1-P(G_2)} \\
&= \frac{\frac{1}{3}\times\frac{3}{10}\times\frac{7}{9} + \frac{1}{3}\times\frac{7}{15}\times\frac{8}{14} + \frac{1}{3}\times\frac{5}{25}\times\frac{20}{24}}{1-\frac{29}{90}} = \frac{20}{61},
\end{aligned}
$$

其中$P(G_2)$利用了抽签模型与次序无关的性质，所以$P(G_2)=P(G_1)$。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：

$$
\begin{aligned}
P(G_1|B_2) &= \sum_{i=1}^{3} P(D_i|B_2)P(G_1|D_iB_2) \\
&= P(D_1|B_2)\frac{3}{10-1} + P(D_2|B_2)\frac{7}{15-1} + P(D_3|B_2)\frac{5}{25-1}
\end{aligned}
$$

其中

$$
P(D_1|B_2) = \frac{P(D_1)P(B_2|D_1)}{\sum_{i=1}^{3} P(D_i)P(B_2|D_i)} = \frac{\frac{1}{3}\times\frac{7}{10}}{\frac{1}{3}\times\frac{10-3}{10} + \frac{1}{3}\times\frac{15-7}{15} + \frac{1}{3}\times\frac{25-5}{25}} = \frac{21}{61},
$$

同理可得

$$
P(D_2|B_2) = \frac{16}{61}, \qquad P(D_3|B_2) = \frac{24}{61}.
$$

代入第一个等式，得到

$$P(G_1|B_2) = \frac{20}{61}.$$

这里在计算$P(G_1|B_2)$时，我们使用了条件概率形式的全概率公式。
警告提示：有些同学会错误地认为$P(D_i|B_2) = P(D_i)$，而忽视了已经发生的事件对选择地区时概率的影响。

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

$$P(B_2) = P(B_1) = 1 - \frac{3+7+5}{10+15+25} = \frac{7}{10}.$$

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$C_1 \cup C_2 \cup C_3$，

$$\begin{aligned} P((C_1 \cup C_2 \cup C_3)G_1|B_2) &= \frac{P((C_1 \cup C_2 \cup C_3)G_1B_2)}{P(B_2)} \\ &= \frac{\sum_{i=1}^{3} P(C_iG_1B_2)}{P(B_2)} \\ &= \frac{\frac{3\times7}{50\times49} + \frac{7\times8}{50\times49} + \frac{5\times20}{50\times49}}{\frac{7}{10}} = \frac{177}{1715}. \end{aligned}$$

题3 袋中有a只黑球，b只白球，把球随机地一只只摸出来，求第k次摸出黑球的概率。$(1 \le k \le a+b)$

解法1： 把a只黑球和b只白球都看作不同的，如果把摸出的球依次放在排列成一直线的$a+b$个位置上作为一个样本点，则样本空间为$a+b$只球的全排列，共有$|\Omega| = (a+b)!$个这样的排列，而且所有的排列是等可能的。有利事件样本点数：在第k个位置上放置一个黑球有a种方式，另外的位置排列剩余的球有$(a+b-1)!$种方式。故所求概率为:

$$\frac{a(a+b-1)!}{(a+b)!} = \frac{a}{a+b}.$$

解法2： 把a只黑球看成没有区别，b只白球也看成没有区别，这时的样本空间是从$a+b$个不同位置中挑出a个位置的不同组合，共有$|\Omega| = C_{a+b}^{a}$个这样的组合，所有的组合是等可能的（为什么？）。有利事件样本点数：由于第k个位置已经选出，故在余下的$a+b-1$个位置中选择$a-1$个有C_{a+b-1}^{a-1}种。故所求概率为:

$$\frac{C_{a+b-1}^{a-1}}{C_{a+b}^{a}} = \frac{a}{a+b}.$$

解法3： 将$a+b$只球逐一编号，样本空间为第k次摸球的全部可能结果，这样的样本有$|\Omega| = a+b$个，每个样本是等可能的（为什么？）。有利样本点为a个，所求概率为$\frac{a}{a+b}$。

解法4： 记$A_k(a,b)$表示"从最初有a只黑球和b只白球的袋中不放回取球第k次取到黑球"。则$P(A_1(a,b))=\frac{a}{a+b}$，当$k>1$时，用"首步分析法"（即用过程的第一步的不同结果作划分，使用全概率公式）得到递推关系：

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

这里的关键想法是：**将已知发生的事件造成的影响，等效于不同初始条件下新的取球过程，从而相应的条件概率等于新过程中的（无条件）概率。**

利用这个（三个参数的）递推关系以及边界条件

$$P(A_k(0,b))=0,\quad P(A_k(a,0))=1,\quad \forall k,a,b$$

和初始条件

$$P(A_1(a,b))=\frac{a}{a+b},\quad \forall a,b$$

使用数学归纳法，得到

$$P(A_k(a,b))=\frac{a}{a+b}.$$

这个解法相对于此前各种解法的优点在于，无须作过多的等可能假设，而且建立递归关系的基本想法很简单。当然，会带来一定技术上的困难，比如数学归纳法等（但是，数学归纳法有它自身的优势，因为它可以把结论以归纳假设形式转化为已知条件，因此结论越强，归纳假设就越强，对你的论证就越有利）。这个解法适用于更一般的取球模型（比如Polya坛子模型，教材第49页习题1.4.26）。

解法5： 用"末步分析法"（即用过程中当前时刻的上一次的不同结果作划分，使用全概率公式）：

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

对$P(A_{k-1})$和$P(\overline{A_{k-1}})$，我们可以使用归纳假设。对条件概率$P(A_k|A_{k-1})$和$P(A_k|\overline{A_{k-1}})$，我们作以下分析。

$$P(A_k|A_{k-1})=\frac{[a+b-(k-2)]P(A_{k-1})-1}{a+b-(k-1)},$$

这里分母$a+b-(k-1)$是第k次取球前（即经过$k-1$次取球后）袋中的总球数，分子中的$a+b-(k-2)$是第$k-1$次取球前（即经过$k-2$次取球后）袋中的总球数，它与第$k-1$次取到黑球的概率$P(A_{k-1})$的乘积相当于第$k-1$次取球前袋中的黑球数，由于已知第$k-1$次取到了一个黑球，所以分子$[a+b-(k-2)]P(A_{k-1})-1$就是第k次取球前袋中的黑球数。类似地，我们可以得到

$$P(A_k|\overline{A_{k-1}})=\frac{[a+b-(k-2)]P(A_{k-1})}{a+b-(k-1)}.$$

将这两个条件概率的表达式代入上面的全概率公式，得到递推公式

$$\begin{aligned}P(A_k) &= P(A_{k-1})\frac{[a+b-(k-2)]P(A_{k-1})-1}{a+b-(k-1)}+(1-P(A_{k-1}))\frac{[a+b-(k-2)]P(A_{k-1})}{a+b-(k-1)}\\ &= P(A_{k-1}).\end{aligned}$$

于是

$$P(A_k)=P(A_1)=\frac{a}{a+b}.$$

如果你有耐心，我们回过头来再审视一下上面那两个条件概率，用最后的结果代入得到

$$P(A_k|A_{k-1}) = \frac{[a+b-(k-2)]\frac{a}{a+b}-1}{a+b-(k-1)},$$

这时，你不难发现分子通常不是整数，因此它不是第k次取球前袋中黑球的个数，那么先前那番解释是不是有问题呢？
如果你还有更多的耐心，我们来计算一下$P(A_3|A_2)$的值，

$$\begin{aligned}P(A_3|A_2) &= \frac{P(A_2A_3)}{P(A_2)} = \frac{P(A_1A_2A_3)+P(\overline{A_1}A_2A_3)}{P(A_1A_2)+P(\overline{A_1}A_2)}\\ &= \frac{\frac{a}{a+b}\times\frac{a-1}{a+b-1}\times\frac{a-2}{a+b-2}+\frac{b}{a+b}\times\frac{a}{a+b-1}\times\frac{a-1}{a+b-2}}{\frac{a}{a+b}\times\frac{a-1}{a+b-1}+\frac{b}{a+b}\times\frac{a}{a+b-1}}\\ &= \frac{a-1}{a+b-1},\end{aligned}$$

而此前的解释给出的值为

$$\frac{(a+b-1)\frac{a}{a+b}-1}{a+b-2} = \frac{(a+b)(a-1)-a}{(a+b)(a+b-2)}.$$

这说明关于条件概率的那番看似合理的解释实际上是站不住脚的。

解法6：

$$\begin{aligned}P(A_2) &= P(A_1)P(A_2|A_1)+P(\overline{A_1})P(A_2|\overline{A_1})\\ &= \frac{a}{a+b}\times\frac{a-1}{a+b-1}+\frac{b}{a+b}\times\frac{a}{a+b-1} = \frac{a}{a+b}.\end{aligned}$$

对$k>2$，记

$$m = \max(0, k-2-b), \qquad M = \min(a, k-2),$$

对$m \le j \le M$，记B_j表示“前$k-2$次取到j个黑球”。则

$$P(A_{k-1}) = \sum_{j=m}^{M} P(B_j)P(A_{k-1}|B_j) = \sum_{j=m}^{M} P(B_j)\frac{a-j}{a+b-(k-2)}.$$

另外，

$$\begin{aligned}&P(A_{k-1}|B_j)P(A_k|A_{k-1}B_j)+P(\overline{A_{k-1}}|B_j)P(A_k|\overline{A_{k-1}}B_j)\\ =&\frac{a-j}{a+b-(k-2)}\times\frac{a-j-1}{a+b-(k-1)}+\frac{b-(k-2-j)}{a+b-(k-2)}\times\frac{a-j}{a+b-(k-1)}\\ =&\frac{a-j}{a+b-(k-2)},\end{aligned}$$

因此

$$\begin{aligned}P(A_k) &= \sum_{j=m}^{M} P(B_j)\left[P(A_{k-1}|B_j)P(A_k|A_{k-1}B_j)+P(\overline{A_{k-1}}|B_j)P(A_k|\overline{A_{k-1}}B_j)\right]\\ &= \sum_{j=m}^{M} P(B_j)\frac{a-j}{a+b-(k-2)}\\ &= P(A_{k-1}).\end{aligned}$$

因此$P(A_k) = P(A_2) = P(A_1) = \frac{a}{a+b}$。

注： 本题结果再次说明：抽签对参与者是公平的，中签的概率与抽签的顺序无关。

题4 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，问第k次摸球时，摸到黑球的概率是多少？

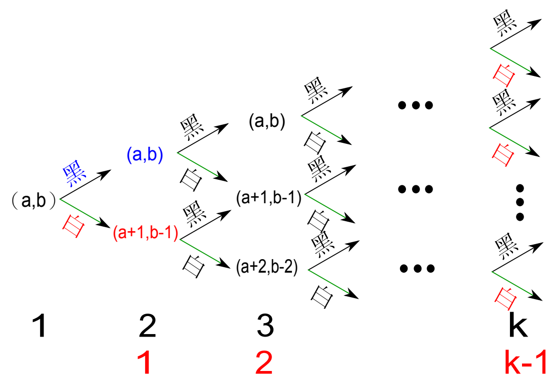
解法1： 记$A_k(a,b)$表示"从最初有a个黑球和b个白球的袋中按上述规则取球，第k次取到黑球"，用首步分析法：

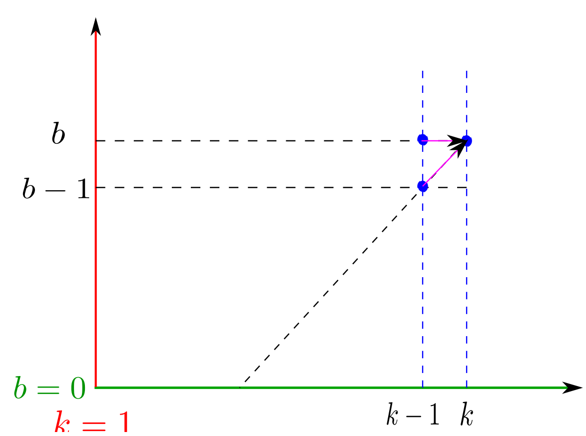
$$
\begin{aligned}
&P(\overline{A_k(a,b)})\\
=&P(A_1(a,b))P(\overline{A_k(a,b)}|A_1(a,b)) + P(\overline{A_1(a,b)})P(\overline{A_k(a,b)}|\overline{A_1(a,b)})\\
=&\frac{a}{a+b}P(\overline{A_{k-1}(a,b)}) + \frac{b}{a+b}P(\overline{A_{k-1}(a+1,b-1)}) \qquad \text{（为什么？）}
\end{aligned}
$$

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: 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

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

然后再此基础上用数学归纳法可以证明

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

在这个解法中，两个条件概率值的确定是至关重要的。为了说明上述直观说法的合理性，有学生给出了如下证明：

首先，

$$P(A_2|A_1)=\frac{a}{a+b}=P(A_1),\qquad P(A_2|A_1^c)=\frac{a+1}{a+b}=\frac{a}{a+b}+\frac{1}{a+b}=P(A_1)+\frac{1}{a+b}.$$

对$k>2$，记p_m为前$k-2$次共取出m个白球的概率，$0\le m\le\min(b,k-2)$。于是

$$P(A_{k-1})=\sum_{m=0}^{\min(b,k-2)}p_m\times\frac{a+m}{a+b}.$$

另一方面，

$$\begin{aligned}
P(A_k|A_{k-1})&=\sum_{m=0}^{\min(b,k-2)}p_m\times\frac{a+m}{a+b}=P(A_{k-1}),\\
P(A_k|A_{k-1}^c)&=\sum_{m=0}^{\min(b,k-2)}p_m\times\frac{a+m+1}{a+b}\\
&=\sum_{m=0}^{\min(b,k-2)}p_m\times\frac{a+m}{a+b}+\frac{1}{a+b}\sum_{m=0}^{\min(b,k-2)}p_m\\
&=P(A_{k-1})+\frac{1}{a+b}.
\end{aligned}$$

不少同学不住地点头道："这个解法和解释应该对。"但实际上，**这个解法真不对！！！为什么呢？**

事实上，

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

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

所以

$$P(A_3|A_2) = \frac{a^3+b(a+1)^2}{[a^2+b(a+1)]\,(a+b)} \neq P(A_2).$$

那么上面那个证明哪里有问题呢？让我们仔细分析一下，记B_m表示事件"前$k-2$次共取出m个白球"。于是

$$P(A_k|A_{k-1}) = \sum_{m=0}^{\min(b,k-2)} P(B_m|A_{k-1})P(A_k|B_mA_{k-1}),$$

上面的证明错误认为$P(B_m|A_{k-1}) = P(B_m)$，而这是不对的。在仅有b个白球的情况下，如果第$k-1$次取到了白球，那么此前取到m个白球的概率就会降低，反之亦然，一个极端情形是，如果前$k-2$次已经取光了所有白球，那么第$k-1$次就必然只能取到黑球。因此A_{k-1}和B_m是不独立的。

虽然直觉经常帮助我们做出发现，但其隐含的跳跃式思维过程中有些细节往往是重要而不能忽视的，这个例子说明，在概率问题的讨论中仅靠直觉是远远不成的，甚至我们会被一些貌似合理的直觉所误导。另外，导致这个错误的一个原因就是初学概率论的学生尚未养成良好的书写习惯和思维习惯，他们往往在尚未理清事件之间的关系和影响之前就急于从数值的角度尽快得到问题的答案。再有，就是初学者对条件概率的认识还只停留在已知发生的事件对试验结果（样本）造成的影响，而忽略了对可能性大小的作用。

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"。

解法3 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，

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

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。