1.3 古典概型

确定概率的古典方法是概率论历史上最早研究的情形，它简单、直观、不需做重复试验.我们先看一个简单试验：掷一颗骰子的试验有6个基本结果.如果六个面是均匀的，那么没有理由认为某个面朝上的可能性会比其他任何一个面朝上的可能性大,可以认为6个基本结果的概率相等，这样的试验称为古典概型。在古典概型中，事件的概率能容易地确定.

如果试验具有下列特点:

（1）试验的基本结果只有有限个，即试验的样本空间为有限样本空间；

（2）所有基本事件发生的可能性相等.

我们称这样的试验为古典概型或等可能概型.

设试验为古典概型，其样本空间含有个样本点，由等可能性知



由概率的性质得

,

故

.

更一般地,假设试验的事件包含个样本点，则事件的概率为

.

这便是古典概型中确定概率的计算公式.由此公式确定的概率叫做古典概率.

例1.3.1 将一硬币抛3次，假设每次抛掷中出现正面和反面的可能性相等，求恰好出现1次正面的概率。

解：由于每次抛掷中出现正面和反面的可能性相等，因此该试验的等可能的基本结果有8个.样本空间取为

,

而事件“恰好出现1次正面”包含3个样本点“HTT,THT,TTH”,即,由古典概率的计算公式,所求的概率为

.

注意,在此试验中,若我们以正面出现的次数为试验的基本结果,即样本空间取为,那么“恰好出现1次正面” 包含中的一个样本点,但并不能由此得出的概率为,因为样本空间中的各个基本事件不具有等可能性.可见用古典方法确定概率时一定要注意:基本事件的等可能性.

当样本空间中样本点较多时,我们不必将样本点一一列举出来,而只需求出样本空间中样本点总数和事件包含的样本点个数.和的计算需要用到计数原理,计数问题往往并不简单,这里我们主要介绍古典概型中几种常见的模型.

一、抽样模型

例1.2.2(不放回抽样模型)假设个元素由两类元素组成, 其中有个A类元素,个B类元素,从这个元素中任取个元素,求事件“取出的个元素中有个A类元素”()的概率.

解:样本空间中样本点的总数为,且这个基本事件是等可能的,而包含的样本点个数为.故所求的概率是

.

注意以上模型的各种变化，看下面例题.

例1.2.3 一个袋中有7个球,其中5个红球,2个白球.从中依次取3次球,每次取一个,取后不放回.试求下列事件的概率,

(1) “前2次取到红球,最后1次取到白球”;

(2) “取到2个红球,1个白球”;

(3) “第2次取到红球”.

解: (1).

(2) .

或 

(3) .

或.

从问题（2）的两种解法可以看出，古典概率的计算需要应用计数原理，但两者还是有区别的，概率计算问题中，重要的不是采用何种计数模式，而是要保证两点：（1）样本空间和随机事件中的样本点的个数要在同一计数模式下计算；（2）样本空间中的各个样本点(即基本事件)出现的概率相同.从问题(3)可以看出,第2次取到红球的概率与第1次取到红球的概率相等,更一般地若7个人依次各取1球,那么每人取到红球的的概率都是.这就是“抽签与顺序无关”的原理,从概率论上解释了“抓阄的公平性”.

以上模型可推广至个元素由多类元素组成的情形,比如,一批产品共件,其中有件一等品, 件二等品, 件三等品(),从中任取件产品,则取到件一等品, 件二等品, 件三等品()的概率为

.

例1.2.4(放回抽样) 个元素由两类元素组成,其中有个A类元素,个B类元素,从中依次取次,每次取一个元素,取后放回,再取下一个.求事件“取出的个元素中有个A类元素”的概率.

解: 由于每次抽取都是从个元素中随机取一个,因此样本空间中样本点的总数为,并且这个基本事件是等可能的.事件包含的样本点个数为.故所求的概率是

.

将上面结果变形为

,

其中,这是什么模型的概率问题(这是你们中学学过的)?

此模型也可推广至个元素由多类元素组成的情形.

二、 盒子模型,或放球模型

例1.2.5 设有个球,每个球都等可能地放到个盒子中的任一个,盒子的容量不限.

(1)设,求球放入不同盒子中的概率;

(2)求某指定的盒子为空盒的概率;

(3)求某指定的盒子有个球的概率.

解:由于每个球都是等可能地放入任一盒子中,因此此试验共有个等可能的基本结果.

(1)设表示事件“球放入不同盒子中”,则包含有个样本点,从而

.

(2) 设表示事件“某指定的盒子为空盒”,则包含有个样本点,从而

.

(3) 设表示事件“某指定的盒子有个球”,则包含有个样本点,从而

.

以上结果可变形为



由变形后的结果会想到什么概率模型?

许多表面上提法不同的概率问题,如果能归于某一模型,那么可用相应的方法去求解.比如历史上一个颇为著名的问题—生日问题就属盒子模型的问题.

假设每个人的生日在一年365天中的任一天是等可能的,随机选取个人,这个人的生日各不相同的概率是多少？在此问题中，365天相当于盒子模型中365个盒子，因此个人的生日各不相同的概率为

.

从而可得到个人中至少有两人的生日相同的概率为

.

下表给出了取不同的具体值所对应的至少有两人生日相同的概率:

|  |  |
| --- | --- |
|  |  |
|  |  |

从上表可以看出, 时,至少有两人的生日相同的概率就超过了50%,而64个人中至少有两人生日相同的概率就很接近于1.初看到这个结果是不是有点吃惊?同学们不妨实地做一下调查.

利用概率的性质,我们可以求解一些较复杂的古典概率的计算问题.

例1.3.3 口袋中装有编号为…的个球，从中有放回地任取次，每次取1个球，求取出的球的最大编号为的概率。

若不放回地任取个球，那么这个问题是简单的，在时,所求的概率为

.

对于有放回地取球，直接考虑事件“取出的球的最大编号为”会比较复杂.如果我们设法把事件表示为某些事件的运算,且这些事件的概率容易求得,那么可使问题变得简单.下面给出解答.

解 考虑事件“取出的球的最大编号不超过”，,事件发生当且仅当每次取出的球的编号为中某一个.因此

.

又因为，且，从而得

.

例1.3.4(配对问题) 在一个有个人参加的晚会上，每人各带了一件礼物.晚会期间各人从放在一起的礼物中随机地取一件，问:(1)至少有一人取到自己的礼物的概率是多少？(2)恰有一人取到自己的礼物的概率是多少?

(1)表示事件“第个人取到自己的礼物”，.那么事件 “至少有一个人取到自己的礼物”可表示为.由于

，，

， ，

……

,

利用概率的加法公式可得

.

(2) 表示事件“第个人取到自己的礼物,而其余的都没有取到自己的礼物”，.那么事件 “恰有一人取到自己的礼物”可表示为.利用(1)的结论可得

，

,

又由于互不相容，利用有限可加性,所求的概率为

.

当较大时,上面两个概率值的计算都比较麻烦.我们来考虑当较大时这两个的概率的近似值.利用,可得,.当较大时，至少有一人取到自己的礼物的概率的极限为，也就是说当较大时至少有一个人取到自己带的礼物的概率的约等于.这表明：即使参加晚会的人很多，至少有一个人取到自己的礼物的概率也不会接近于1.同样地有.

1.3.2 几何方法

古典概率适合的随机试验有两个特点：（1）样本空间为有限样本空间，（2）一切基本事件的地位均等.在实际中，我们还会遇到这样一类随机现象：所有基本事件的地位均等,但基本结果有无穷多个.例如，从1,2,3,4,5,6中任取一个数，考查取到的数不小于4的概率.这是古典概率的问题.如果把该问题改为；从区间中任取一个数，考虑取到的数不小于4的概率，这个试验的所有基本事件的地位均等,其样本空间为区间，有无穷多个基本事件.容易想到这个试验中的事件（即样本空间的子集）的概率应与事件的长度成正比,这便是几何概率.

如果试验的样本空间为度量有限的区间或区域（这里所说的度量是指区间的长度，区域的面积、体积等），且样本点在样本空间中均匀分布（样本点在样本空间中均匀分布的详细内容是：样本点落在中任一有度量的子集中的可能性大小只与子集的度量成正比），这样的试验称为几何概型.

设试验为几何概型，样本空间为，为一事件，则事件的概率为

,

其中,分别为和的度量.这样确定的概率叫做几何概率.

例1.2.6 （1）从区间中任取一个数，求取到的数不小于4的概率；

（2）从区间中任取两个数，求取到的两数之和不小于8的概率.

解 (1)试验的样本空间为,设事件“取到的数不小于4”记为，那么，该试验是几何概型，由几何概率的计算公式得

.

（2）从区间中任取的两个数分别记为，试验样本空间为,设事件“取到的两数之和不小于8”记为，那么 ，由几何概率的计算公式得

.

思考题,补充及应用问题

思考题

1.将6个球随机地放入5个盒子中,5个盒子分别编号1,2,3,4,5.

(1)1号,2号,3号,4号,5号盒子中分别有2个球,2球,1个球,1个球,和0个球的概率是多少?

(2)6个球的分布状况是2,2,1,1,0(即有2个盒子各放2个球,3个盒子各放1个球,1个盒子没有球)的概率是多少?

2. 将6本书随机地分发给三位同学(每本书分发各位同学的可能性相同),问三位同学各分得2本书的概率是多少?

3.若将上面问题改为:将6本书(其中3本数学书,3本语文书)随机地平均分给3位同学(每位同学2本书),求3位同学各分得一本数学书、一本语文书的概率?

4.10对夫妇随意地坐成一圈,10对夫妇都相邻而坐的概率是多少?没有一对夫妻坐在一起的概率是多少?( ,)

补充

1.确定概率的主观方法

在实现世界中，有许多随机现象是不可能在相同条件下重复的，这时有关事件的概率如何确定呢？这种情形下，事件的概率无法用频率去解释.如果把不可重复的随机现象排除在概率论的研究对象之外,那么概率论的理论与方法的应用将受到很大限制.为了突破这种限制,人们提出了确定概率的主观方法.

统计学界有两个学派:频率学派和贝叶斯学派.频率学派主张概率的频率解释.而贝叶斯学派认为:一个事件的概率是人们根据经验对该事件发生的可能性所给出的个人信念.这样给出的概率称为主观概率.

利用经验确定事件发生可能性大小的例子很多.

(1)在气象预报中,说“明天下雨的概率为90%”，这是气象部门根据近期的气象条件及专业知识给出的判断.站在给定的时间点上考察明天是否下雨,这是无法做到在相同条件下重复的.这个概率是主观概率.人们听到这个信息后会做好防雨准备.

(2)一个企业家根据认为“某种新产品在未来市场上畅销”的概率为80%，这是企业家依据他或他的团队多年的经验和市场信息等作出的判断，也是这位企业家对这个事件的确信程度.

（3）一位医生根据多年的临床经验和一位患者的病情,判断“这位病人治愈”的可能性有90%。

从以上例子可以看出：

（1）主观概率并不是主观臆造出来的，需要当事人对所考察的的事件有丰富的知识和经验，并对历史信息和当时的信息进行仔细考察和分析，最后确定对事件发生的确信程度.这样确定的主观概率是可信的.

(2)用主观方法确定的事件的概率，本质上是根据已有的信息对随机事件概率的一种估计.虽然结论的准确性有待实践去检验和修正,但结论的可信性在统计意义上是有价值的.

(3)对于无法大量重复的随机现象,谨慎地给出主观概率并指导决策和行动是有意义的,久这一点看，主观概率至少是频率方法的一种补充.

用主观方法确定的概率也要符合概率的公理化定义.这样主观概率可完全纳入概率论的理论框架中.

2.贝特朗悖论

几何概率在现代概率概念的发展中曾经起过重大作用.19世纪时,不少人相信，只要找到适当的等可能性描述，就可以给概率问题以唯一的解答，然而有人构造出这样的例子，它包含着几种似乎都有道理但却相互矛盾的答案，下面就是一个著名的例子.

(贝特朗悖论)在半经为1的圆内随机地取一条弦,问其长超过该圆内等边三角形边长的概率为多少?

这是一个几何概率的问题,但是基于对术语“随机地”的含义的不同解释，这个问题却有多种不同的答案，下而是其中三种.

解法1 任何弦交圆周两点.不失一般性,先固定其中一点(记为)于圆周上，以此点为顶点作等边三角形，显然弦长超过该圆内等过三角形边长当且仅当弦的另一个端点落在点的三角形对边所对应的优弧上,而该优弦的长度为圆周长的,故所求的概率为.

解法2 弦的长度取决于圆心到弦的距离,而与弦的方向无关,因此可以假定它垂直于某一直径.当且仅当弦与圆心的距离小于时其长大于,故所求的概率为.

解法3 弦长被其中点的位置唯一确定, 当且仅当其中点位于半径为的同心圆内时弦长大于,故所求的概率为.

同一个问题有三种不同的答案，细究其原因，发现是取弦时采用了不同的等可能性假定.在解法1中,假设端点在圆周上均匀分布;解法2则假定弦的中点在直经上均匀分布;而解法3又假定弦的中点在圆内均匀分布.实际上,这三个答案是针对三种不同的随机试验，对于各自的随机试验而言,它们都是正确的。

在使用术语“随机”，“等可能”，“均匀分布”等时，应明确其含义。

1899年贝特朗在巴黎出版的《概率论》一书中对几何概率提出了批评，并以生动的实例引起大家的注意。这种批评推动了概率论的发展。

由于采用等可能性来定义概率这一基本概念有这种困难，促使人们思考该如何给出概率的定义，因此后来就选择另外的途经，即在定义概率的概念时只指明概率应具用的基本性质，而把具体概率的确定放在一边.这就出现了我们前面已经介绍的概率的公理化定义,这样做的好处就是概率论可以包含各种各样的随机现象.

应用问题

3.抽样模型在抽样调查中是很重要的.比如件产品中有件不合格品, 件合格品,从中抽检件产品.在实际问题中,的值已知,而的值未知,要根据抽查的件产品中的次品件数去推断的值,这属于统计问题.也有的值已知,而的值未知的情况,例如，从某一湖里,捕捉尾鱼,将这些鱼作上某种标记,然后放回湖中,过一段时间后,再从湖中捕捞出尾鱼,发现其中有尾有标记,对该湖中鱼的总数能作出什么结论?我们可以这样考虑这个问题. 从湖中捕捞出尾鱼,其中有尾有标记的概率为

,

这里是已知的,因此上述概率依赖于,我们的问题是取多大时,这个概率最大?

4.（蒲丰投针问题）平面上画有间隔为的等距平行线,向平面任意投掷一枚长为()的针,求针与任一平行线相交的概率.

首先考虑如何刻画针的位置,以及找出针与平行线相交的充要条件的公式化表示.针的中点位置及针与平行线的夹角可以刻画针的位置,为此引入两个变量:表示针的中点与最近的平行线的距离;表示针与该平行线的夹角.显然,. “向平面内任意投针”的概率意义是:在区间上取任意值是“等可能”的, 在区间上取任意值也是“等可能”的,从而在平面区域上取任意点是“等可能”的.因此可得样本空间为平面上的矩形区域:

，

而针与平行线相交的充要条件是,这样事件“针与平行线相交”可表示为样本空间的子集:



从而



如果的值给定,并用的值代入上式可算得.反之,如果知道的值以及的值便可求出的值.那么的值如何知道呢?我们可以多次重复这个试验,以获得的频率去近似的值,从而得到的近似值.历史上确有一些学者做过这个试验,下表记录了他们的试验结果:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 试验者 | 年份 |  | 投掷次数 | 相交次数 | 的近似值 |
| Wolf  Fox  Lazzerini  Reina | 1850  1884  1901  1925 | 0.8  0.75  0.83  0.54 | 5000  1030  3408  2520 | 2532  489  1808  859 | 3.1596  3.1595  3.1415  3.1795 |

这是一个很神奇的想法:设计一个随机试验,使一个事件的概率与某个未知量有关,然后通过大量重复试验以获得这个事件的频率,再以频率估计概率,从而求出未知量的近似值.一般说来,试验次数越多近似精度越高.靠人工难以大量重复试验和保证准确,现在可把这个工作交给计算机去完成.这种方法现已得到了迅速发展和广泛应用,以至于成为了标准计算方法之一,称这种方法为随机模拟法,也称为Monte Carlo法.当然这里还有许多问题需要解决,比如,如何产生随机数?如何设计有效的算法?精度如何?达到要求的精度应模拟多少次?等等.

蒲丰于1877年所提出的投针试验常被认为是蒙特卡罗方法的起源. 蒙特卡罗方法也称为统计模拟方法,是由冯.诺依曼在20世纪40年代中期为研制核武器而首次提出的.在第二次世界大战中蒙特卡罗曾被用作绝密计算方法的代号,该计算的目的是预测原子弹中的中子通量,因数以百万计的中子,沿着大量铀分子中的随机通道辐射的情况只能在计算机上模拟而无法进行理论上的预测.蒙特卡罗是摩纳哥的著名赌城, 蒙特卡罗方法借用该城市的名称来象征性地表明其特点:一是由于通道的变化是随机的;二是生产原子弹具有高度风险性,本身就是一次大赌博

5(游程问题)论下面讨论一下游程问题，游程理论可以有很多方式应用于统计中，主要用于随机性检验和齐次性检验.

在由两类元素构成的有序序列中，由同一类元素组成的（没有异类元素隔开）最大的子序列称为一个游程（或一个连贯）。例如，6个和4个排成序列 ，那么该序列有3个游程，个游程，总游程数为.

设由个元素和个元素排成序列，总游程数记为，若总游程数过小则说明同类元素有聚集现象，若过大则说明有混杂现象。为了从统计上侦别是否有某种系统因素造成以上现象，为此需计算在这（）个元素随机排成一列的条件下，游程数取各个可能值的概率。即计算.

首先可以看出这个元素（同类元素不加区分）排成一列时，有（或）个可以区分的排列结果，在这个元素随机排列的条件下，这个结果是等概的.

下面考虑事件包含的基本结果数.若（即为偶数），则游程数和游程数均为，游程数为的排列结果数为（它等于方程的正整数解的个数）。同样地，游程数为的排列结果数为，前面的游程可以是游程也可以是游程。所以事件“总游程为”包含了个基本结果。故

，

同样地可得

.

注意以上计算结果中我们都约定：若或，.

游程理论有很多应用,比如休哈特把游程理论应用于工业生产的质量控制中，当橡皮垫圈制成后，其厚度各不相同，厚垫圈的长串(长游程)的出现，可能反映生产过程不够完善,由此引导我们去发现不完善的原因.再比如,在农作物的实验中，有病作物的长串提示疾病的蔓延.等等.

6（计票问题）候选人得到张选票而候选人得到张选票，现在一张一张地计票，假定选票的一切排列次序都是等可能的，(1) 若,求在计票过程中的票数始终领先的概率;(2)若,求在计票过程中的票数始终不落后于的概率.

引入变量

，，，

，。

画出()的散点图,并在相邻两点用线段连接形成一条折线,称为一条路经.路经下两个端点是固定的,左端点为,右端点为,且由条件知每条路经的概率相等.全部的路经数为.先介绍下面引理.

引理（反映原理） 从点到点的路经中触到或者穿过轴的路经数等于从点到点的路经数，这里为整数，点为点关于轴的对称点.

证明 对于任意一条点到点且触到或者穿过轴的路经，其与轴的第一个交点记为，可构造这样一条从点到点的路经：从点到点的一段和点到点的一段关于轴的对称，而从点到点的一段与原来相同.反之，对任意一条点到点的路经,也可类似地构造一条从一条点到点且触到或者穿过轴的路经.可见从点到点的路经中触到或者穿过轴的路经与从点到点的路经具有一一对应的关系.因此两者的路经数相同.

注: 从点到点的路经数等于,其中.

从点到点的路经中触到或者穿过轴的路经数等于从点到点的路经数等于,其中.

推论1 从点到点的路经中位于轴上方的路经数等于.

证明 从点到点的路经中触到或者穿过轴的路经数等于为.因此从点到点的路经总数为,因此从点到点的路经中位于轴上方的路经数为

.

由推论1 ,可得的票数始终领先的概率为

.

推论2 从点到点的路经中不穿过轴的路经数等于

证明 不穿过轴的路经等同于位于直线上方的路经数.因此从点到点的路经中不穿过轴的路经数等于从点到点的路经中位于轴上方的路经数.而后者的路经数为

.

由推论1 ,可得的票数始终不落后于的概率为

.

特别当时, .例如，一天中有50人到银行柜台存钱,50人取钱,假设都存1元钱或取1元钱.那么柜台服务员在这一天中不用使用银行备用的钱的概率为.