四、事件之间的关系

**4.0变量关系**

数据科学中的许多问题都是围绕核心问题，就是关于变量之间关系。

* 如何使用属性来判断细胞是否是癌性的？
* 加利福尼亚县平均收入与用水之间的关系是什么？
* 汽车购买者支付好的里程还是快速加速？您在数据8中遇到了许多其他类似问题。

关于随机变量之间关系，概率论有助于我们提出并回答精确性问题。特别是，在给定另一组的值情况下，它帮助我们理解另外一组随机变量的条件行为。

在本章中，我们将研究定义在同一组输出结果上的多重随机变量。

**联合分布**

假设X和Y是定义在同一输出空间上的两个随机变量。We will use the

记号p（X= x，Y= y），表示X具有x和Y具有y值的概率。

X和Y的联合分布由所有概率 P(X=x,Y=y)组成，其中（x，y）分布在所有（x，y）（X，Y）的可能值范围内。

例如：

在三次掷硬币的中，让X成为头两次抛头中的头数，Y是最后两次抛头中的头数。那么

P(X=0,Y=2)=0

P(X=1,Y=1)=P(THT or HTH)=2/8

所有其他概率都是1/8，你可以通过检查三次抛掷的六个剩余结果来判断。例如，

P(X=1,Y=2)=P(THH)=1/8

对X和Y的约束条件是，每一个都必须在{01,1,2}的范围内，并且 |x−y|<2。让我们定义一个函数，它以X和Y作为参数，并返回p（X= x，Y= y）。

In [12]:

**def** joint\_probability(x, y):

**if** x == 1 & y == 1:

**return** 2/8

**elif** abs(x - y) < 2:

**return** 1/8

**else**:

**return** 0

.

PROF140库包含用于显示两个随机变量的联合分布的表格方法。作为第一步，需要两个变量中的每一个变量的可能值。在我们的示例中，两个值都有{01,1,2}，因此相同的列表或数组可以同时表示两个变量。

In [3]:

k = np.arange(3)

为了构造一个联合分布对象，我们必须首先构造所有可能的值对和概率的表。调用方法为:

Table().values(variable\_name\_1, values\_1, variable\_name\_2, values\_2).probability\_function(function\_name)

其中函数名 function\_name是以xx和yy为参数并返回p（X= x，Y= y）的函数。In [13]:

joint\_table = Table().values('X', k, 'Y', k).probability\_function(joint\_probability)

joint\_table

Out[13]:

| **X** | **Y** | **概率** |
| --- | --- | --- |
| 0 | 0 | 0.125 |
| 0 | 1 | 0.125 |
| 0 | 2 | 0 |
| 1 | 0 | 0.125 |
| 1 | 1 | 0.25 |
| 1 | 2 | 0.125 |
| 2 | 0 | 0 |
| 2 | 1 | 0.125 |
| 2 | 2 | 0.125 |

此表显示联合分布，为了检验这确实是一个分布，我们可以把所有的概率加起来。总和是1，因此它应该是一个分布。

In [14]:

joint\_table.column(2).sum()

Out[14]:

1.0

**4.1联合分布表**

虽然上述表格数据，确实显示了联合分布，但以不同的方式以更传统、更据启发性地观察相同信息。

方法 to\_joint将上述表转换为一个“联合分布对象”，它显示为X和Y的常规联合分布表。

In [15]:

joint\_dist = joint\_table.to\_joint()

joint\_dist

Out[15]:

|  | **X=0** | **X=1** | **X=2** |
| --- | --- | --- | --- |
| **Y=2** | 0.000 | 0.125 | 0.125 |
| **Y=1** | 0.125 | 0.250 | 0.125 |
| **Y=0** | 0.125 | 0.125 | 0.000 |

Each cell corresponds to a pair (x,y)(x,y), where x is a value of X and y a value of Y. In the cell you see P(X=x,Y=y)P(X=x,Y=y), the probability of the pair (x,y)(x,y).

每个单元对应于一对（x，y），其中x是X的值，y为Y的值。在单元格中，你记p（x= x，y= y），为点对（x，y）的概率。

类似于在数据8中看到的列联表，联合分布表使用在分析两个分类变量之间的关系。在列联表中，每个单元包含在一对特定类别中的个体数。在联合分布表中，如上面的一个表，每个单元包含一对特定值的概率。

**概率计算**

比如,

该表包含关于X和Y之间关系的完整信息。为了找出X和Y所确定的任何事件的概率，简单地识别使事件发生的每个单元，并累加他们的频率。通过划分事件，我们应用这种基本方法来发现概率。

比如:

P(X>Y)=P(X=1,Y=0)+P(X=2,Y=0)+P(X=2,Y=1)

=0.125+0+0.125=0.25

以及

P(X=Y)=P(X=0,Y=0)+P(X=1,Y=1)+P(X=2,Y=2)

=0.125+0.25+0.125=0.5

**4.2边际分布**

联合分布表根据对（x，y）的值来划分整个结果空间。

在我们的例子中，我们掷硬币三次。XX是头两次投掷的次数，YY是最后两次投掷中的头数。这里又是XX和YY的联合分布表。

In [3]:

joint\_dist

Out[3]:

|  | **X=0** | **X=1** | **X=2** |
| --- | --- | --- | --- |
| **Y=2** | 0.000 | 0.125 | 0.125 |
| **Y=1** | 0.125 | 0.250 | 0.125 |
| **Y=0** | 0.125 | 0.125 | 0.000 |

这个空间被分割成9块，每个都有自己的机会。所有的机会总数是1，正如我们在前一节中看到的。

**根据Y分区{X= x}**

现在看标记为x＝0的列。在该列的每个单元格中，x的值为0，y是Y范围内的一些元素。因此，列x＝0根据y的值对事件{x＝0 }进行分区，并显示每个分区的概率。

事实上，对于每一个X，

{X=x}=⋃all y{X=x,Y=y}

,

这是一个两两不相交的集合。因此，通过加法规则，

P(X=x)=∑all yP(X=x,Y=y)

也就是说，p（X= x）是列X= x中概率的和。因为p（X= x）是X分布中的一般项，我们已经知道，我们可以从X和Y的联合分布中导出XX的分布。

为了找到X分布的数值，我们将使用一种称为“边际”的方法，该方法以X作为参数，并使用在联合分布对象上。当我们看到输出时，使用“边际”这个词的原因就会变得清晰。

In [4]:

joint\_dist.marginal('X')

Out[4]:

|  | **X=0** | **X=1** | **X=2** |
| --- | --- | --- | --- |
| **Y=2** | 0.000 | 0.125 | 0.125 |
| **Y=1** | 0.125 | 0.250 | 0.125 |
| **Y=0** | 0.125 | 0.125 | 0.000 |
| **总和: X的边际** | 0.250 | 0.500 | 0.250 |

现在在表的底部，所有的列和，这构成了X的分布。这些总和出现在表的边际处，因此这种分布称为边际分布。

这只是X的概率分布的一个新名称，头两次抛硬币的数量。

**两个边际**

对于变量X，能做什么？你也可以沿着行看Y。

In [8]:

joint\_dist.marginal('Y')

Out[8]:

|  | **X=0** | **X=1** | **X=2** | **总和: Y的边际** |
| --- | --- | --- | --- | --- |
| **Y=2** | 0.000 | 0.125 | 0.125 | 0.25 |
| **Y=1** | 0.125 | 0.250 | 0.125 | 0.50 |
| **Y=0** | 0.125 | 0.125 | 0.000 | 0.25 |

Y也是硬币投掷两次的正面向上的次数（三次投掷的最后两次），所以边际分布的概率是有意义的。

你也可以同时得到两个边缘：

In [5]:

joint\_dist.both\_marginals()

Out[5]:

|  | **X=0** | **X=1** | **X=2** | **总和: Y的边际** |
| --- | --- | --- | --- | --- |
| **Y=2** | 0.000 | 0.125 | 0.125 | 0.25 |
| **Y=1** | 0.125 | 0.250 | 0.125 | 0.50 |
| **Y=0** | 0.125 | 0.125 | 0.000 | 0.25 |
| **总和：X的边际** | 0.250 | 0.500 | 0.250 | 1.00 |
|  |  |  |  |  |

右下角单元格是表中所有概率的总和，也是每个边距中所有概率的总和。令人欣慰的是，它是1。

比较两个边缘分布，它们是相同的（可能的值和相应的概率）。所以X自定义相等于Y。

**4.3条件分布**

要理解两个变量之间的关系，必须检查每个变量的条件行为，并给出另一个变量的值。为了实现这一目标，我们将首先检查前面的章节的简单例子，然后发展一般理论。

在我们的例子中，三个掷硬币的正面向上的次数，其中X是头两次抛硬币的正面向上的次数，Y是最后两次抛硬币的正面向上的次数，X和Y的联合分布和两个边缘显示在下表中。

In [7]:

joint\_dist.both\_marginals()

Out[7]:

|  | **X=0** | **X=1** | **X=2** | **Sum: Marginal of Y** |
| --- | --- | --- | --- | --- |
| **Y=2** | 0.000 | 0.125 | 0.125 | 0.25 |
| **Y=1** | 0.125 | 0.250 | 0.125 | 0.50 |
| **Y=0** | 0.125 | 0.125 | 0.000 | 0.25 |
| **总和：X的边际** | 0.250 | 0.500 | 0.250 | 1.00 |

给定变量X的一个特定值x，也就是说，在前两次抛掷硬币有x次正面向上，我们可以算出最后两次抛掷硬币的正面向上次数的条件分布。

在随机变量语言中，对于变量X的每个值x，随机变量Y在条件X＝x下有分布情况。我们可以得到这些三个条件的所有分布如下：

In [8]:

*# conditional distribution of Y given each different value of X*

*#* 给定x情况下，每个不同值的y的条件分布

joint\_dist.conditional\_dist('Y', 'X')

Out[8]:

|  | **Dist. of Y | X=0** | **Dist. of Y | X=1** | **Dist. of Y | X=2** | **边际分布 of Y** |
| --- | --- | --- | --- | --- |
| **Y=2** | 0.0 | 0.25 | 0.5 | 0.25 |
| **Y=1** | 0.5 | 0.50 | 0.5 | 0.50 |
| **Y=0** | 0.5 | 0.25 | 0.0 | 0.25 |
| **Sum** | 1.0 | 1.00 | 1.0 | 1.00 |

若要理解此表，请从第一列开始。在该列中，给定的条件是x= 0，也就是说，在头两次抛硬币，没有正面向上发生。在这种情况下，Y不可能是2，这就是为什么你看到顶部单元中的概率为0。

如果X＝0，根据第三次掷币是正面还是反面，那么Y只能是1或0。这解释了为什么剩下的两个单元中概率都为0.5。

你可以用这种方法算出所有其他概率。但是你不必使用最初的结果，来计算出这些条件概率。在示例开始，只使用联合分布表和除法规则。

例如，

P(Y=1∣X=0)=P(X=0,Y=1)P(X=0)=0.125/0.25=0.5

很容易看出为什么条件分布表中的每个列总和为1。使用联合分布表中取对应的单元格，将其与其列的总和相除，计算出每列中的每个单元格。所以结果表中的列和1。

**理论知识**

现在我们将概括上面的例子中所做的计算。

让X和Y是定义在同一空间上的两个随机变量。如果x是变量X的一个可能值，y是变量Y的可能值，那么

P(Y=y∣X=x)=P(X=x,Y=y)P(X=x)

因此，对于变量X，给与固定值，则Y的条件分布，是在给定条件x= x \* 下的概率集合。

P(Y=y∣X=)=P(X=,Y=y)P(X=)

y的取值包括在Y的所有值范围内。请记住，Y表示变量的值,是所观察到特定X的值，它是常数。

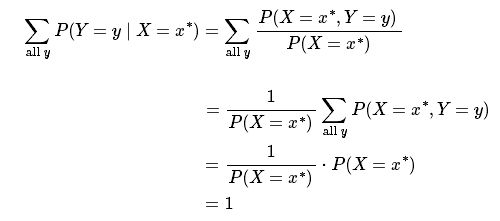
**和为1的条件分布概率**

在一个分布中，概率和必须为1。要验证对于上述定义的条件分布，这也是成立的，需要从基本规则开始。

根据y值根据事件{x= }划分，计算出p（x= x \*）



在给定条件X= x \*下，现在我们将给出Y的条件分布的概率和，求和是1。



**4.4 更新分布**

正如我们在研究贝叶斯规则时所看到的，条件为我们提供了一种基于新数据来更新我们的观点的方法。让我们来看看，基于数据情况下，条件分布如何被用来表示随机参数。

**一个随机抽取的硬币**

假设我有一枚硬币和三个有偏差的硬币。假设每一个有偏差的硬币投掷出正面的概率为0.9。我随意挑选一枚硬币，掷两次，告诉投掷出正面的总数。关于硬币是正常的还是有偏差的，通过两次抛掷出正面向上总数，你能得出什么结论？

在公开正面向上总数前，你的观点应该是，硬币正常公正的机会是0.25，硬币存在偏差的几率是0.75。

这个例子的目的是看关于抛币结果信息是如何影响这个评估观点的。

让R是硬币正面向上的随机概率。R的可能值为0.5和0.9，R的先验概率分布由下表给出。

In [5]:

coins = [0.5, 0.9]

prior = [0.25, 0.75]

Table().values(coins).probability(prior)

Out[5]:

| **Value** | **Probability** |
| --- | --- |
| 0.5 | 0.25 |
| 0.9 | 0.75 |

让H成为两个投掷中的正面向上的次数。然后，R和H的联合分布由下面方式组成。

P(R=r,H=h)  where r∈{0.5,0.9} and h∈{0,1,2}

有六个这样的两两组合，让我们算出其中的一对。使用乘法法则，

P(R=0.9,H=2)=P(R=0.9)P(H=2R=0.9)

=0.75\*0.92=0.6075

事件{H＝1 }发生，如果有一次正面向上，后面一次正面向下，然后又是一个正面向上。所以

P(R=0.5,H=1)=P(R=0.5)P(H=1R=0.5)

=0.25[(0.5\*0.5)+(0.5\*0.5)]

=0.125

依据所有R值和H值，使用相同的推理，我们可以来计算的p（r＝r，h＝h）。

让我们直接在Python中做这件事。函数joint\_probs 以R和H为参数，并返回概率值P（R＝R，H＝H）。

In [7]:

**def** joint\_probs(r, h):

*"""Return P(R = r, H = h)"""*

*# Start with the distribution of the number of heads in two tosses*

*# of a coin that lands heads with a known chance r;*

*# these are the chances of h=0, h=1, and h=2*

heads\_2\_tosses = make\_array((1-r)\*\*2, 2\*r\*(1-r), r\*\*2)

**if** r == 0.5:

**return** 0.25\*heads\_2\_tosses.item(h)

**elif** r == 0.9:

**return** 0.75\*heads\_2\_tosses.item(h)

我们现在可以使用PROF140方法来构造R和H的联合分布表，如联合分布部分所述。回想一下，我们使用硬币和先验知识之前构建变量R的先验分布：

In [8]:

coins, prior

Out[8]:

([0.5, 0.9], [0.25, 0.75])

In [9]:

heads = np.arange(3)

joint\_tbl = Table().values('R', coins, 'H', heads).probability\_function(joint\_probs)

joint\_dist = joint\_tbl.to\_joint()

joint\_dist

Out[9]:

|  | **R=0.5** | **R=0.9** |
| --- | --- | --- |
| **H=2** | 0.0625 | 0.6075 |
| **H=1** | 0.1250 | 0.1350 |
| **H=0** | 0.0625 | 0.0075 |

p（r＝0.9，h＝2）和p（r＝0.5，h＝1））的值与我们直接计算的值一致。

让我们检查一下R的边际分布与从一枚硬币和三个有偏差硬币中随机抽取的分布是否一致。没有正面向上数量的信息，R的分布应该仅仅是先验的。这是，正如你从下表看到的底部边际分布。

In [11]:

joint\_dist.marginal('R')

Out[11]:

|  | **R=0.5** | **R=0.9** |
| --- | --- | --- |
| **H=2** | 0.0625 | 0.6075 |
| **H=1** | 0.1250 | 0.1350 |
| **H=0** | 0.0625 | 0.0075 |
| **总和：r的边际** | 0.2500 | 0.7500 |

现在假设我捡起硬币（秘密地），掷两次，告诉你脑正面向上的次数。通过给定H的值，你如何估计R的值？

一个好方法是在开始阶段，通过给定H的值，找出R的条件分布。这里是所有这些条件分布，对于不同的给定值的H值。

In [13]:

joint\_dist.conditional\_dist('R', 'H')

Out[13]:

|  | **R=0.5** | **R=0.9** | **Sum** |
| --- | --- | --- | --- |
| **Dist. of R | H=2** | 0.093284 | 0.906716 | 1.0 |
| **Dist. of R | H=1** | 0.480769 | 0.519231 | 1.0 |
| **Dist. of R | H=0** | 0.892857 | 0.107143 | 1.0 |
| **R的边缘** | 0.250000 | 0.750000 | 1.0 |

从底向上读取此表。记住硬币是随机抽取的。

如果我不告诉你正面朝上的次数，你可以告诉我关于硬币有25%的可能性是公平的，75%的可能性是一个有偏差的。

如果我告诉你正面朝上的次数为0，你对硬币的看法发生了巨大变化，判读更有利于正常公平的硬币。有偏见的硬币有很小的机会正面朝上，所以即使其中一个很可能被选中，数据倾斜更有利于公平硬币。

如果我告诉你正面朝上的次数为1，你就有点左右为难了。有偏差见的硬币有一个适度的机会（18%）使得正面向上，公平硬币有50%的机会正面向上。但是，抛出的硬币有75%的几率是偏差的，而25%的机会是公平的。这两种效应的大小使你很难确定硬币的类型。你只是略微倾向于“偏差”。

如果我告诉你正面朝上的次数为2，你的观点会明显倾向于偏差硬币。不仅从概率上很容易两次正面向上，而且很可能选择偏差硬币抛掷。

**基于数据更新你的观点**

这是机器学习中经常出现的一个简单例子。

你先从一个关于未知量的先验知识开始。在这个例子中，根据以前的分布知识，公平硬币将被选中的机会25%。

对于未知量的每一个值，数据都有可能性机率。对于我们四个硬币中的每一个，我们知道只要抛出硬币，就有可能得到相应正面向上数量的概率。

当你看到数据后，你对未知量的看法可能会改变，有时改变非常大。这种变化某种可能性方式取决于先验知识。

然后你可以投掷更多，并每次根据新投掷硬币正面向上的次数更新你的判断。

**4.5依赖与独立**

条件分布帮助我们对两个随机变量是否相互独立的直觉概念进行形式化。设X和Y为两个随机变量，假设我们给出了X的值。这会改变我们对Y的看法吗？如果答案是肯定的，那么我们会说X和Y是依赖的。如果答案是否定的，不管X的给定值是什么，那么我们将说X和Y是独立的。

让我们从一些例子开始，然后转到精确的定义和结果。

**依赖**

这里是两个随机变量X和Y的联合分布。由此，我们可以说X和Y是独立的还是依赖的？

In [46]:

dist1

Out[46]:

|  | **X=0** | **X=1** | **X=2** | **X=3** |
| --- | --- | --- | --- | --- |
| **Y=3** | 0.037037 | 0.000000 | 0.000000 | 0.00000 |
| **Y=2** | 0.166667 | 0.055556 | 0.000000 | 0.00000 |
| **Y=1** | 0.250000 | 0.166667 | 0.027778 | 0.00000 |
| **Y=0** | 0.125000 | 0.125000 | 0.041667 | 0.00463 |

你可以立刻看到，如果X= 3，那么Y只能是0，而如果X＝2，那么y可以是1或2。知道X的价值会改变Y的分布。这就是依赖。

这里有一个例子，通过查看可能的值，您无法快速确定依赖性或独立性。

In [22]:

dist2

Out[22]:

|  | **X=3** | **X=4** |
| --- | --- | --- |
| **Y=7** | 0.3 | 0.1 |
| **Y=6** | 0.2 | 0.2 |
| **Y=5** | 0.1 | 0.1 |

但是你可以通过观察在给定XX的条件下,Y的条件分布来判断----它们是不同的。

In [23]:

dist2.conditional\_dist('Y', 'X')

Out[23]:

|  | **Dist. of Y | X=3** | **Dist. of Y | X=4** | **Marginal of Y** |
| --- | --- | --- | --- |
| **Y=7** | 0.500000 | 0.25 | 0.4 |
| **Y=6** | 0.333333 | 0.50 | 0.4 |
| **Y=5** | 0.166667 | 0.25 | 0.2 |
| **Sum** | 1.000000 | 1.00 | 1.0 |

它遵循（并且你应该尝试证明），至少有一些x在边际上条件分布在给定给出不同的Y值也将彼此不同。

注意，并非所有的条件分布都不同。给定y＝5的条件分布与y＝6的条件分布相同。但当y＝7时，条件分布发生变化。X和Y是依赖的。

In [24]:

dist2.conditional\_dist('X', 'Y')

Out[24]:

|  | **X=3** | **X=4** | **Sum** |
| --- | --- | --- | --- |
| **Dist. of X | Y=7** | 0.75 | 0.25 | 1.0 |
| **Dist. of X | Y=6** | 0.50 | 0.50 | 1.0 |
| **Dist. of X | Y=5** | 0.50 | 0.50 | 1.0 |
| **X的边际** | 0.60 | 0.40 | 1.0 |

**独立性**

这里有一个联合分布表，你不能立即判断是否存在依赖关系。

In [7]:

dist3

Out[7]:

|  | **X=0** | **X=1** | **X=2** | **X=3** |
| --- | --- | --- | --- | --- |
| **Y=4** | 0.000096 | 0.000289 | 0.000289 | 0.000096 |
| **Y=3** | 0.001929 | 0.005787 | 0.005787 | 0.001929 |
| **Y=2** | 0.014468 | 0.043403 | 0.043403 | 0.014468 |
| **Y=1** | 0.048225 | 0.144676 | 0.144676 | 0.048225 |
| **Y=0** | 0.060282 | 0.180845 | 0.180845 | 0.060282 |

但是在条件X情况下，你看看在Y的概率分布情况

In [10]:

dist3.conditional\_dist('X', 'Y')

Out[10]:

|  | **X=0** | **X=1** | **X=2** | **X=3** | **Sum** |
| --- | --- | --- | --- | --- | --- |
| **Dist. of X | Y=4** | 0.125 | 0.375 | 0.375 | 0.125 | 1.0 |
| **Dist. of X | Y=3** | 0.125 | 0.375 | 0.375 | 0.125 | 1.0 |
| **Dist. of X | Y=2** | 0.125 | 0.375 | 0.375 | 0.125 | 1.0 |
| **Dist. of X | Y=1** | 0.125 | 0.375 | 0.375 | 0.125 | 1.0 |
| **Dist. of X | Y=0** | 0.125 | 0.375 | 0.375 | 0.125 | 1.0 |
| **Marginal of X** | 0.125 | 0.375 | 0.375 | 0.125 | 1.0 |

所有的行都是一样的。也就是说，给定不同Y的值，X的所有条件分布都是相同的，因此X的边际分布也相同。

考虑到Y的值，X的概率根本不变。这就是独立性。

你可以通过在条件X上调节Y得出相同的结论。

In [11]:

dist3.conditional\_dist('Y', 'X')

Out[11]:

|  | **Dist. of Y | X=0** | **Dist. of Y | X=1** | **Dist. of Y | X=2** | **Dist. of Y | X=3** | **Marginal of Y** |
| --- | --- | --- | --- | --- | --- |
| **Y=4** | 0.000772 | 0.000772 | 0.000772 | 0.000772 | 0.000772 |
| **Y=3** | 0.015432 | 0.015432 | 0.015432 | 0.015432 | 0.015432 |
| **Y=2** | 0.115741 | 0.115741 | 0.115741 | 0.115741 | 0.115741 |
| **Y=1** | 0.385802 | 0.385802 | 0.385802 | 0.385802 | 0.385802 |
| **Y=0** | 0.482253 | 0.482253 | 0.482253 | 0.482253 | 0.482253 |
| **Sum** | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 |

**两个事件的独立性**

独立的概念似乎是直观的，但是如果不慎重定义，它可能会遇到很多麻烦。所以让我们正式定义它吧。

两个事件的独立性有两个等价的定义。第一个封装了独立的主要思想，第二个是有利于计算。

如果P(B∣A)=P(B)，则两个事件A和B是独立的。

等价地，如果P(AB)=P(A)P(B)，则A和B是独立的。

**两个随机变量的独立**

我们在本节的例子中观察到的，可以转化为对’独立’的正式定义。

如果变量X的每个x值和变量Y的每一个值y，则两个随机变量X和Y是独立的，

P(Y=y∣X=x)=P(Y=y)

也就是说，无论给定X是什么，Y的条件分布与我们知道X=x与否无关。

在事件独立性方面的一个等价定义是，对于X和Y的任何值，事件{X= x}和{Y= y}是独立的。

也就是说， X和Y是独立的，对于任意X的取值和Y的取值。

P(X=x,Y=y) = P(X=x)P(Y=y)

独立性简化了乘法规则中的条件概率。

事实上，如果X和Y是独立的随机变量，那么由X确定的任何事件都独立于Y所确定的任何事件。例如，如果X和Y是独立的，x是一个数，则{X= x}与{Y> x}无关。此外，X的任何函数都独立于Y的任何函数。

你可以通过分区和独立性的定义来证明这些事实。证明是有章可循的，但有些很费力。如果你不想证明，欢迎你接受事实。

**相互独立**

事件A1、A2、…AN是相互独立的（或独立于短），如果任何子集的事件发生，所有其他子集的条件概率保持不变。

那真是太好了。实际上，这意味着不管你知道的事件发生了什么，涉及其他事件的机会是不变的。

在随机变量方面，X1、X2、…、XNX1、X2、…、Xn是独立的，如果给定任何子集的值，则由剩余变量确定的事件的机会不变。

在实践中，这只是形式化的陈述，如“每次投掷硬币的结果是独立的”或“随机抽签与替换是独立的”。

尽量不要拘泥于形式主义。请注意，理论不仅支持直觉，而且强化直觉。你可以期待你的直觉在这门课程结束时比现在更敏锐！

**“I.I.D.”随机变量**

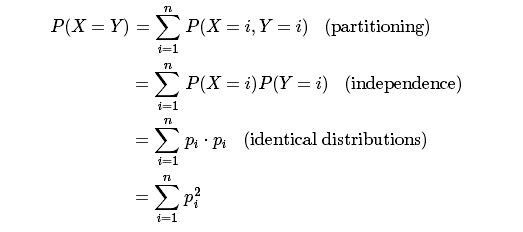
如果随机变量是相互独立和相同分布的，则称为“I.I.D”，这是概率论中最著名的首字母缩略词之一。你可以把I.I.D.随机变量看作是从一个群体中置换出来的，或者作为同一个实验的独立复制的结果。

假设X的分布是由

P(X=i)=pi,   i=1,2,…,n

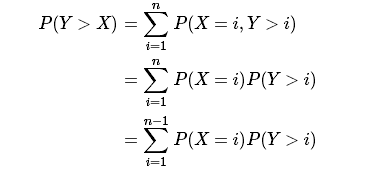
where ∑ni=1pi=1∑i=1npi=1. Now let XX and YY be i.i.d. What is P(X=Y)P(X=Y)? We'll answer this question by using the fundamental method, now in random variable notation.

其中 ∑ni=1pi=1。现在让x和y为i.i.d。什么是P（x= y）？我们将用基本方法来回答这个问题，现在用随机变量表示。



如果你知道所有PIPI的数值，最后一个表达式是容易计算的。

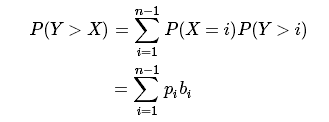
以同样的方式，



因为 P(Y>n)=0，那么对每一i，下式

C:\Users\thinkpad\AppData\Roaming\Tencent\Users\157351451\QQ\WinTemp\RichOle\6JIZ@$FNZK1R$5UA)_UBMT9.png成立。

把 “大于i”的事件求和，记为bi。然后



这也是一个简单的计算，如果你知道所有的pi。对于n=4，它归结为

p1⋅(p2+p3+p4)  +  p2⋅(p3+p4)  +  p3⋅p4