概率论公理

样本空间和事件

考虑一个试验，其结果是不可肯定地预测的。当然，尽管在试验之前无法得知结果，但是假设所有可能结果的集合是已知的。所有可能的结果构成的集合，称为该试验的**样本空间**(Sample Space)，并记为S。

样本空间的任一子集E称为**事件**(Event)，换句话说，事件就是由试验的某些可能结果组成的一个集合。如果试验的结果包含在E里面，那么就称事件E发生了。

对于同一个样本空间S的任意两个事件E和F，定义一个新的事件$E\cup F$，它由以下结果组成：这些结果或在E中，或在F中，或既在E中也在F中。也就是说，如果事件E或者事件F中有一个发生，那么$E\cup F$就发生。事件$E\cup F$称为事件E和事件F的**并**(Union)。类似地，对于任意两个事件E和F，还可以定义事件EF，称为E和F的**交**(Intersection)，它由E和F的公共元素组成。即事件EF(有时也记作$E\cap F$)发生当且仅当E和F同时发生。不可能发生的事件被称为**不可能事件**，记为$\emptyset$（即$\emptyset$是不包含任何结果的事件）。如果$EF=\emptyset$，则称E和F是**互不相容的**(Mutually Exclusive)。

最后，对于任意事件E，可以定义E的补，记为E^c，它表示包含在样本空间中但不包含在E中的所有结果构成的事件。即E^c发生当且仅当E不发生。注意，样本空间S的补$S^c=\emptyset$。

对于任意两个事件E和F，如果E中的所有结果都在F中，那么称E包含于F或E是F的子集，记为$E\subset F$(或者$F\supset E$，有时候称F是E的一个**超集**)。因此，如果$E\subset F$，那么E发生就表示F也发生。如果$E\subset F$并且$F\subset E$，那么就称E和F是相等的，记为$E=F$。

维恩图(Venn Diagram)是一种用来阐述事件之间的逻辑关系的非常实用的集合表示方法。样本空间S用一个大的矩形表示，意即它包含了所有的可能结果。事件$E,F,G,\cdots$用包含在矩形之内的一个个小圆表示，所关心的事件可以用相应的阴影区域表示。

事件的并、交、补遵循类似于代数学里的一些运算法则，例如：

$$
\begin{aligned}
&\text{交换律：}E\cup F=F\cup E && EF=FE\\
&\text{结合律：}(E\cup F)\cup G=E\cup(F\cup G) && (EF)G=E(FG)\\
&\text{分配律：}(E\cup F)G=EG\cup FG && EF\cup G=(E\cup G)(F\cup G)
\end{aligned}
$$

上述关系可以通过以下方式来证明：先证明等式左边的事件中的任一结果必然包含于等式右边的事件中，再证明等式右边的事件中的任一结果也包含于等式左边的事件中，另一种证明方法就是利用维恩图。

下面关于事件的交、并、补这三个基本运算之间的重要的关系式称为**德摩根律**(DeMorgan's Laws)：

$$
\left(\bigcup_{i=1}^{n}E_i\right)^c=\bigcap_{i=1}^{n}E_i^c \qquad \left(\bigcap_{i=1}^{n}E_i\right)^c=\bigcup_{i=1}^{n}E_i^c
$$

例如，对于事件E和F，德摩根律表明$(E\cup F)^c=E^cF^c$以及$(EF)^c=E^c\cup F^c$。对于一般的n，为了证明德摩根律，首先假设x是$\left(\bigcup_{i=1}^n E_i\right)^c$里的一个元素，那么$x\notin\bigcup_{i=1}^n E_i$，这就意味着$x$并不包含于任意一个事件$E_i(i=1,2,\dots,n)$，所以$\forall i(i=1,2,\dots,n),x\in E_i^c$，即$x\in\bigcap_{i=1}^n E_i^c$。另一方面，假设$x\in\bigcap_{i=1}^n E_i^c$，那么$\forall i(i=1,2,\dots,n),x\in E_i^c$，这就意味着$x\notin E_i$，所以$x\notin\bigcup_{i=1}^n E_i$，即$x\in\left(\bigcup_{i=1}^n E_i\right)^c$，这样就证明了德摩根律的第一条。第二条同理可证。

概率论公理

一种定义事件发生概率的方法是利用事件发生的**相对频率**。定义如下：假设有一个样本空间为S的试验，它在相同的条件下可重复进行。对于样本空间S中的事件E，记$n(E)$为n次重复试验中事件E发生的次数。那么，该事件发生的概率$P(E)$就定义如下：

$$
P(E)=\lim_{n\to\infty}\frac{n(E)}{n}
$$

即定义概率$P(E)$为E发生的次数占试验总次数的比例的极限，也即E发生相对频率的极限。不过实际上这个定义是有很严重的缺陷的：我们怎么知道$\frac{n(E)}{n}$会收敛到一个固定的常数？而且如果进行另一次重复的试验，它也会收敛到这个相同的常数？用相对频率来定义概率的支持者常常认为$\frac{n(E)}{n}$趋于某常数极限值是整个系统的一个假设，或者说一个**公理**(Axiom)。

假设某个试验的样本空间为S，对应于其中任一事件E，定义一个数$P(E)$，它满足如下3条公理：

$$0 \le P(E) \le 1$$

$$P(S) = 1$$

$$\text{对于任一列互不相容的事件}E_1, E_2, \cdots (\text{即如果}i \ne j\text{，则}E_iE_j = \emptyset)\text{，有}$$
$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} E_i$$

我们把满足以上三条公理的$P(E)$称为事件E的概率。

公理1说明，任何事件E的概率在0到1之间。公理2说明，S作为必然发生的事件，其概率定义为1。公理3说明对任意一列互不相容事件，至少有一事件发生的概率等于各事件发生的概率之和。

几个简单命题

命题1：

$$P(E^c) = 1 - P(E)$$

该命题说明，一个事件不发生的概率等于1减去它发生的概率。其证明过程非常简单。

证明命题1：

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

命题2：如果$E \subset F$，那么$P(E) \le P(F)$。

证明命题2：因为$E \subset F$，所以可以将F表示为

$$F = E \cup E^c F$$

这样，E和E^cF是互不相容的，由公理3可得：

$$P(F) = P(E) + P(E^cF) \ge P(E)$$

命题3：

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

命题4：容斥恒等式(Inclusion-Exclusion Identity)

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1}E_{i_2}) + \cdots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r}) + \cdots + (-1)^{n+1} P(E_1E_2\cdots E_n)$$

$$\Updownarrow$$

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} \cdots E_{i_r})$$

其中，$\sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})$表示对一切下标集合$\{i_1, i_2, \ldots, i_r\}$所对应的值求和，和项一共包含$\binom{n}{r}$项。换言之，$n$个事件并的概率等于这些事件独自发生的概率之和，减去两个事件同时发生的概率之和，再加上三个事件同时发生的概率之和……

在容斥恒等式中，右边如果只取第一项，那么得到事件并的概率的一个上界，如果只取前两项，那么得到事件并的概率的一个下界。

第一个容斥不等式，即

$$P\left(\bigcup_{i=1}^{n} E_i\right) \le \sum_{i=1}^{n} P(E_i)$$

被称为布尔不等式(Boole's Inequality)

等可能结果的样本空间

在很多试验中，一个很自然的假设是，样本空间中的所有结果发生的可能性都是一样的，即考虑一个试验，其样本空间S是有限集，不妨设为$S = \{1, 2, \ldots, N\}$，那么就经常会自然地假设

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\})$$

$$\Updownarrow$$

$$P(\{i\}) = \frac{1}{N} \qquad i = 1, 2, \ldots, N$$

再利用公理3即可得到，对于任意事件E，

$$P(E) = \frac{E\text{中的结果数}}{S\text{中的结果数}}$$

换言之，如果假定一次试验中的所有结果都是等可能发生的，那么任何事件E发生的概率等于E中所含有的结果数占样本空间中的所有结果数的比例。

概率：连续集函数

事件序列$\{E_n,\ n \ge 1\}$如果满足

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

则称为递增序列，反之，如果满足

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

则称为递减序列。如果$\{E_n,\ n \ge 1\}$是递增事件序列，那么定义一个新的事件，记为$\lim_{n\to\infty} E_n$，如下：

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

类似的，如果$\{E_n,\ n \ge 1\}$是递减事件序列，那么定义$\lim_{n\to\infty} E_n$如下：

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

命题1：如果$\{E_n,\ n \ge 1\}$是递增或者递减事件序列，那么

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

概率：确信程度的度量

我们在前面解释过，一个事件的概率，是指在重复进行某个试验的情况下，对该事件发生频率的一种度量。然而，在其他情况下也会使用术语概率。最自然又简单的解释是，概率是人们对自己的说法的确信程度的一种度量。概率作为个体确信程度的度量经常被称为**主观概率**(Subjective Probability)。无论把概率解释为确信程度的度量，还是事件发生的频率，其数学属性都不会改变。