概率中的PDF、PMF、CDF

一、概念解释

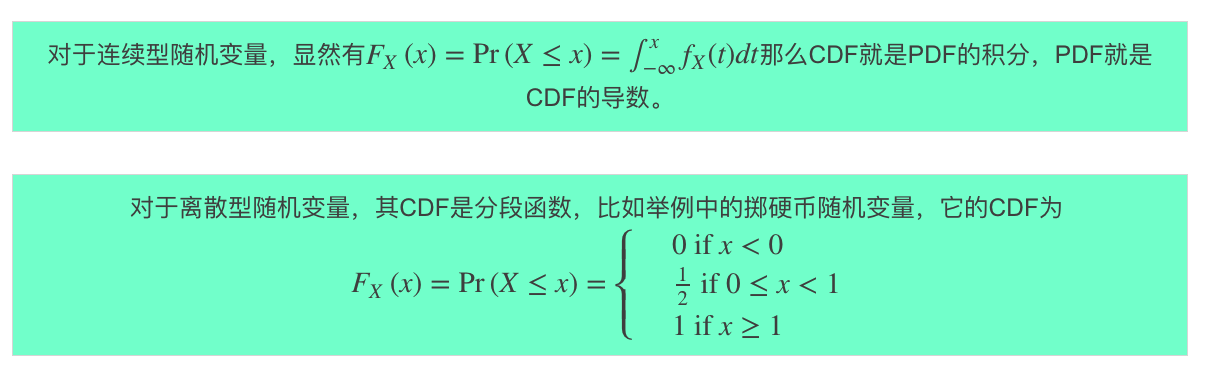
* PDF：概率密度函数(probability density function)，在数学中连续型随机变量的概率密度函数是一个描述这个随机变量的输出值，在某个确定的取值点附近的可能性函数。
* PMF：概率质量函数(probability mass function)，在概率论中，概率质量函数是离散随机变量在各特定取值上的概率。
* CDF：累计分布函数(cumulative distribution function)，又叫分布函数，是概率密度函数的积分，能完整地描述一个实随机变量X的分布概率。

二、数学表示

* PDF：如果X是连续型随机变量。定义概率密度函数为，用PDF在某一区间上的积分来刻画随机变量落在这个区间中的概率，即：
* PMF：如果X离散型随机变量，定义概率质量函数为,PMF其实就是高中所学的离散型随机变量的分布律，即：

比如掷一枚均匀地硬币，如果正面令X=1，如果反面就令X=0，那么它的PMF就是：

* CDF：不管是什么类型（连续、离散、其他）的随机变量，都可以定义为它的累积分布函数，有时简称为分布函数。



三、概念分析

根据上述，我们能得到以下的结论：

* PDF是连续随机变量特有的，PMF是离散型随机变量特有的。
* PDF的取值本身不是概率，它是一种趋势（密度），只有对连续型随机变量的取值进行积分后才是概率，也就是说对于连续值确定它在某一点的概率是没有意义的。
* PMF的取值本身代表该值的概率。

四、分布函数的意义

从两点来分析分布函数的意义：

* 1）为什么需要分布函数：对于离散型随机变量，可以直接用分布律来描述其统计规律性，而对于非离散型的随机变量，如连续型随机变量，我们无法一一列举出随机变量的所有取值，所以他的概率分布不能像随机变量那样进行描述，于是引入PDF，用积分来求随机变量落入某个区间的概率。分布律不能描述连续型随机变量，密度函数不能描述离散型随机变量，因此需要找到一个统一方式描述随机变量统计规律，这就有了分布函数。
* 2）分布函数的意义：分布函数F(x)在点x处的函数值表示X落在区间()内的概率，所以分布函数就是定义域为R的一个普通函数，因此我们可以把概率问题转化为函数问题，从而可以利用普通的函数知识来研究概率问题，增大了概率的研究范围。