基础知识（Author：Black）

随机变量

1.定义

一个随机事件有很多种可能，可以将每种可能映射到数值，这种**映射**就是随机变量。

随机变量的本质是**函数**，函数的自变量的随机过程，因变量是随机过程对应的数值

2.分类

离散随机变量：随机变量的取值是**离散的**并且是**有限个**

连续性随机变量：随机变量的取值是**连续的**并且有**无穷个**

如何描述随机变量的概率分布情况

1.离散型随机变量

a.分布律：可以理解为概率函数，自变量为随机变量的取值，因变量是该种取值的概率

公式形式：

$$P\{X=x_k\}=p_x,k=1,2,3....$$

列表模式：

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p_i	p_1	p_2	$\cdots$	p_n	$\cdots$

b.分布函数：

设X是一个随机变量，x为任意是实数，则随机变量X的分布函数为：

$$F(x)=P\{X\leq x\},-\infty<x<+\infty$$

2.连续随机变量

概率密度函数：

可以参考下如下链接的说明：https://juejin.im/entry/5bdbc266e51d4505113ca8ba

分布函数：

同离散随机变量类似，连续随机变量的分布函数是概率密度f(x)从负无穷到x的积分，如下：

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

二项分布

定义

伯努利试验：一次实验中只有两种结果*A*及$\overline{A}$

二项分布分布律：一次伯努利试验中，事件A出现的概率为*p*，不出现的概率为*q*=1-*p*。则n次试验中A出现*k*次的概率为：

$$P\{X=k\} = C_n^k p^k q^{n-k}$$

$$X \sim b(n,p)$$

二项分布属性：

X 的期望：$\mu = np$

X 的方差：$\sigma^2 = np(1-p)$

X 的标准差：$\sigma = \sqrt{(np(1-p))}$

泊松分布

定义

泊松分布是由二项分布发展出来的，具体过程可以参考https://www.zhihu.com/question/26441147

泊松分布使用条件：

1.在 两个 相同大小/长度的 时间/空间内, 一个事件的发生的概率是相同的

2.事件发生于不发生是相互独立的\不受其他事件的发生或者不发生影响

3.时间发生的概率很小

当概率*p*的数值很小的时候，泊松分布近似等于二项分布

分布率：

$$P\{X=k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0,1,2...$$

其中：λ是 在 一段特定时间/空间内 事件发生的平均值

e 是自然常数 2.71828...

k是事件在一段时间发生的次数

$P\{X = k\}$是该事件在这一段时间/空间发生k次的概率

正态分布

概率密度函数如下：

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

其中μ为期望，σ为标准差

大数定律

大数定律就以严格的数学形式表现了随机现象的一个性质：平稳结果的稳定性（或者说频率的稳定性）

大数定律从理论上解决：用频率近似代替概率的问题,用样本均值近似代替理论均值

中心极限定理

当样本量N逐渐趋于无穷大时，N个抽样样本的均值的频数逐渐趋于正态分布，其对原总体的分布不做任何要求，意味着无论总体是什么分布，其抽样样本的均值的频数的分布都随着抽样数的增多而趋于正态分布。