W1

sample space:S 样本空间，所有有可能的outcome的集合

S是discrete离散的:当他有有限的Outcome或无限的可数的outcome（1,3，,6,8,9，……但总归是可数的）

S是continuous连续的：当它包含有一个实数区间的时候（比如1到2，他中间是可以无限分割的）

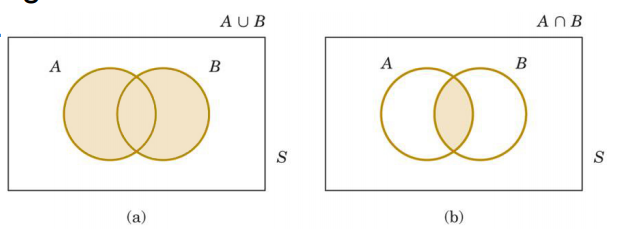
event:E 是S的子集

Union: 

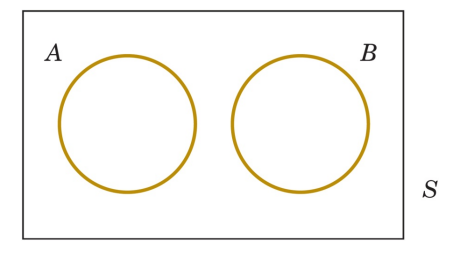
intersection: 

complement: 补集，不属于这个event，但仍是S的所有outcome的子集

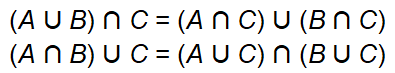
Venn diagrams:文氏图

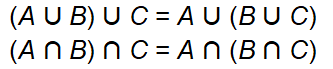


events A与events B是 mutually exclusive关系：没有一个outcome相同，发生了其中一个OUTCOME另外一个就被排除了



Commutative law: 

Distributive law: +

associative law: 

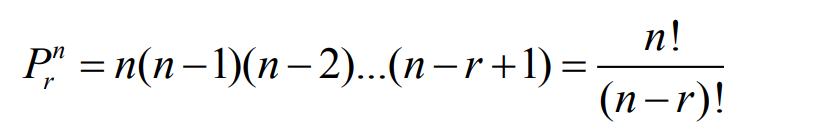
Demorgan’s law: 

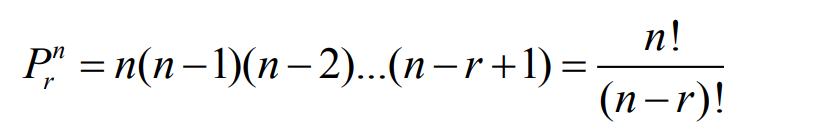
Complement law: 

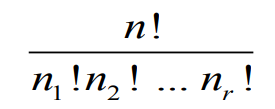
*multiplication rule:n1方法解决step1,* 

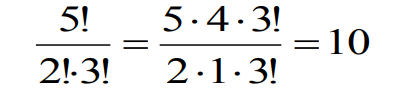
*Permutation排列组合 rule: S = {a, b, c} abc, acb, bac, bca, cab, cba，permutation的数字等于*

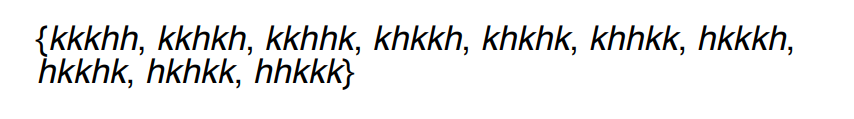
*n个数的集合=n!*

r个子集从n个数

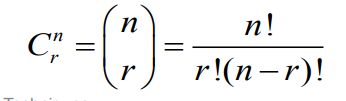
有顺序的不同物体就要permutation，比如8个不同的球里选四个排进4个不同的槽里

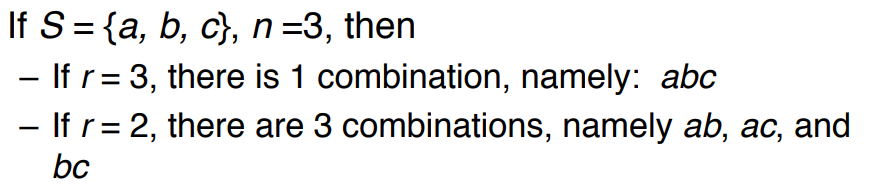
有顺序的相同物体就要，比如需要3个K两个H



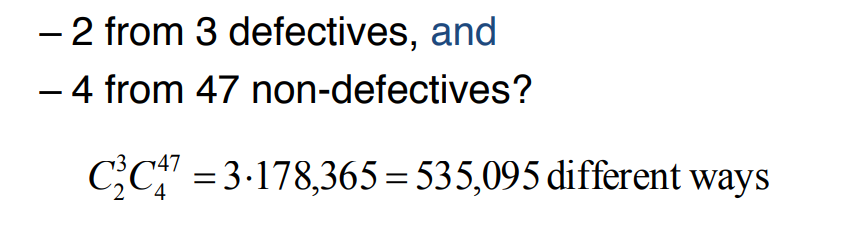


*Combination rule:顺序不重要的时候*

**

例子

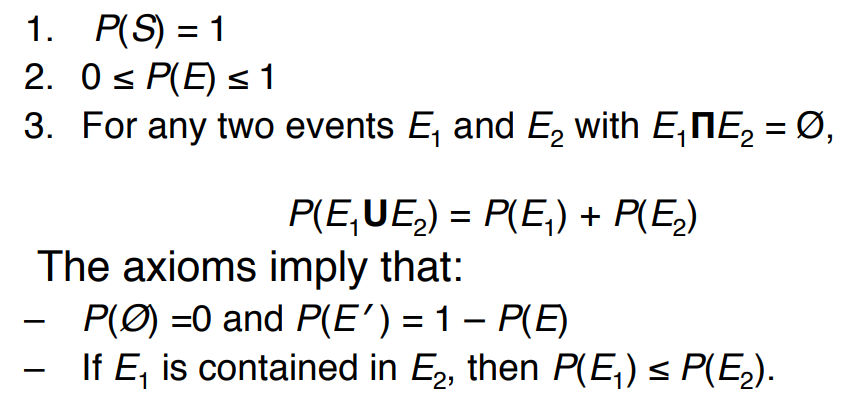
2.50里面有三个defective,47个non-defective,三选2,47选4

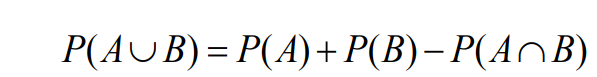


probability：就是在随机的试验中产生一个特定结果的可能性，在之间，1是绝对成立，0是绝对不成立

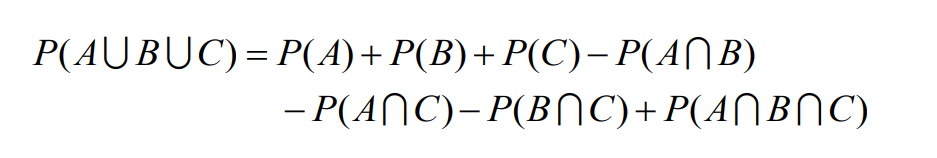
n个等可能的结果，probability就是1/N

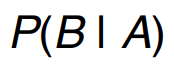
Axioms公理 of Probability：(S是全集，E，是子集)

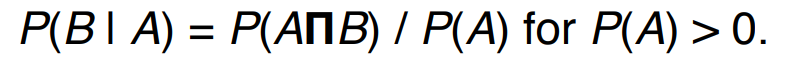


additional rule:如果由两个event

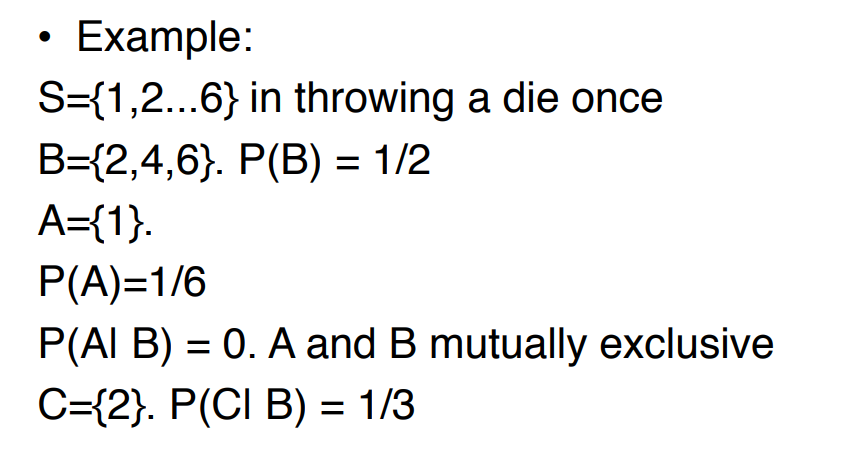
如果两个EVENT完全不相交

三个或三个以上

Conditionally probability: 当事件A已经发生时发生B的可能性。

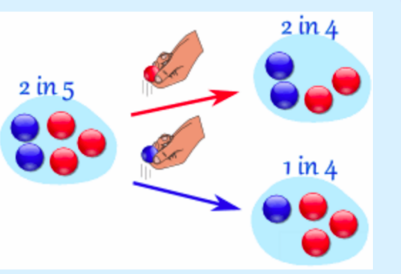


A做前提，发生B的情况=A与B都发生的情况除以A发生的情况

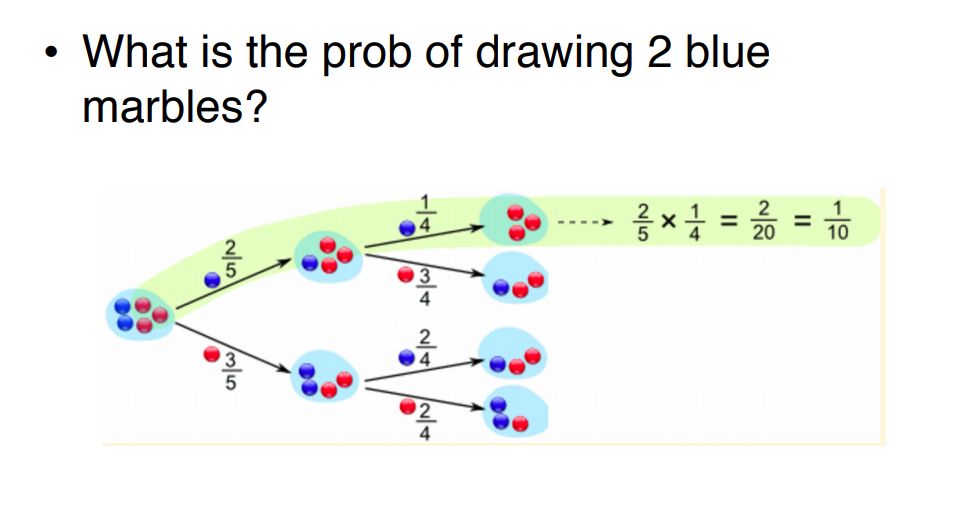


W2

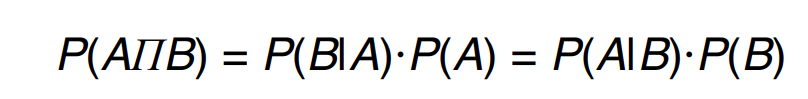
independent event：每个事件相互独立，比如投掷硬币，要么正要么反

dependent event:5个球拿一个蓝的是2in5,但是拿走一个红的和拿走一个蓝的再拿一个蓝的几率是不一样的 

连着拉出两个蓝的情况：

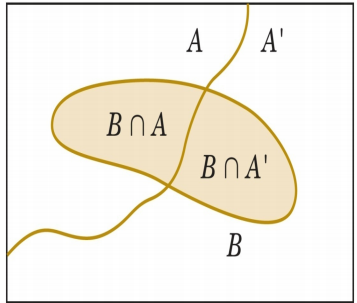


multiplication rule:

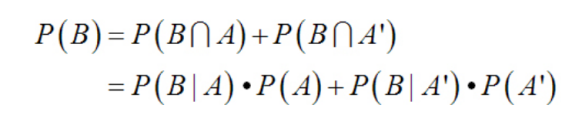


其实不用想这么麻烦，就是发生第一次拿的概率×第二次拿的概率

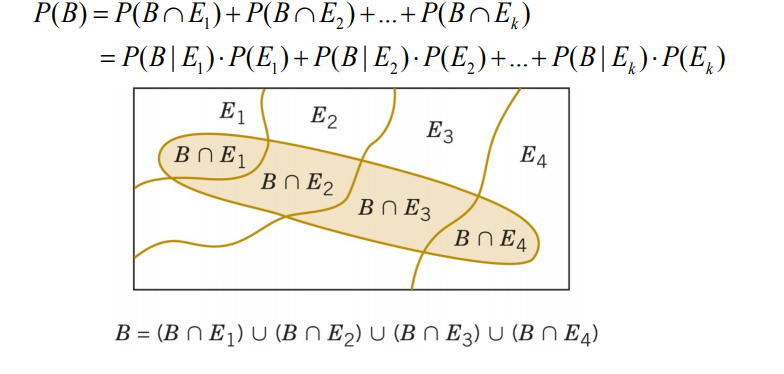
是Mutually exclusive, 也是Mutually exclusive



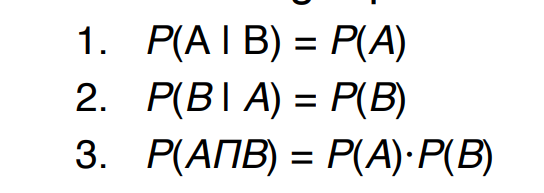
total probability rule



total probability rule在多个情况下

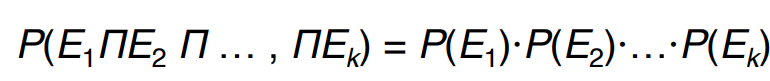


E1 E2…是mutually exclusive的

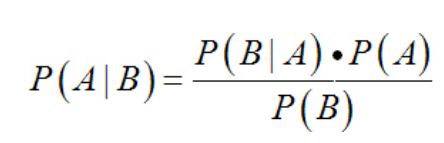


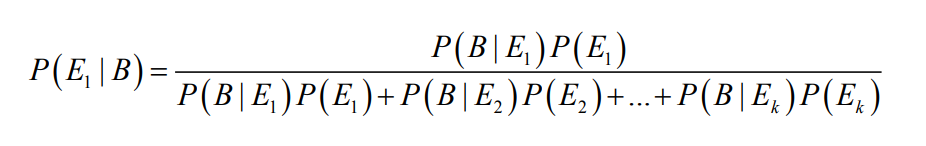
如果下列式子有一个成立，那么A与B就是互相独立的

当有且仅有E1，E2,.,….EK互相独立时



bayes理论

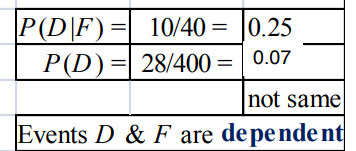


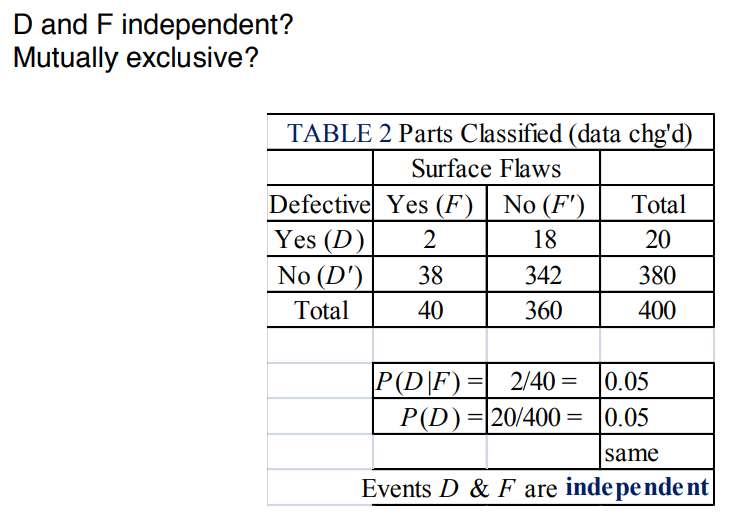


E1，E2，E3，EK相互独立

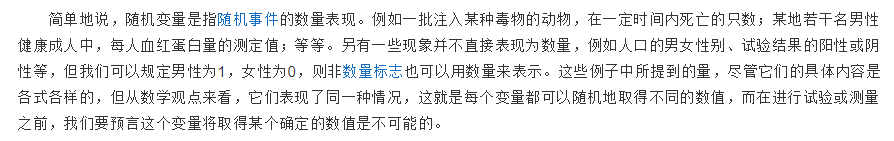
如果两个事件 mutually exclusive，那么几乎必然是非independent的（independent的情况是不管事件A发生没发生，B的几率不变）

F发生以后D的几率变了，DEPENDENT



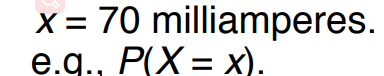


random variable：随机变量，随机变量（random variable）表示随机试验各种结果的实值单值函数。随机事件不论与数量是否直接有关，都可以数量化，即都能用数量化的方式表达。[1]



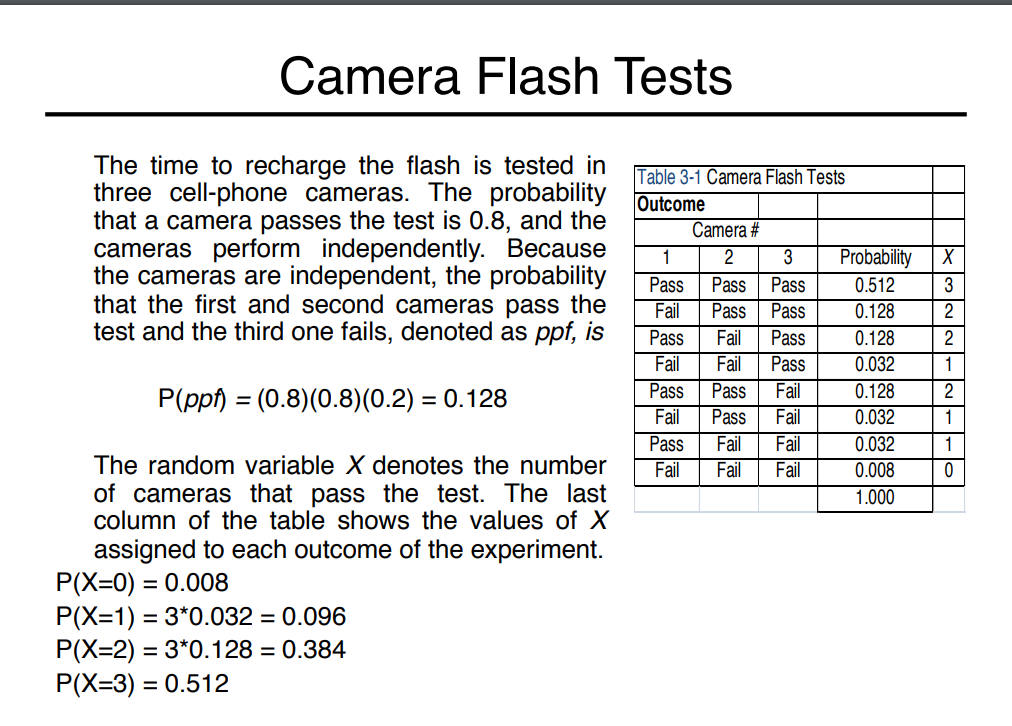
他是每种结果和随机试验总集之间的function,意义在于用可统计的实数表达了关系

随机变量的用大写斜体X表示，测量值用小写斜体x表示

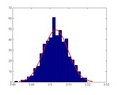
意思就是70毫安的随机变量

discrete random variable:有限Or无限可数，统计方法是数出来的/表面划痕这种

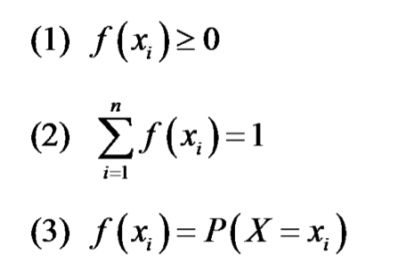
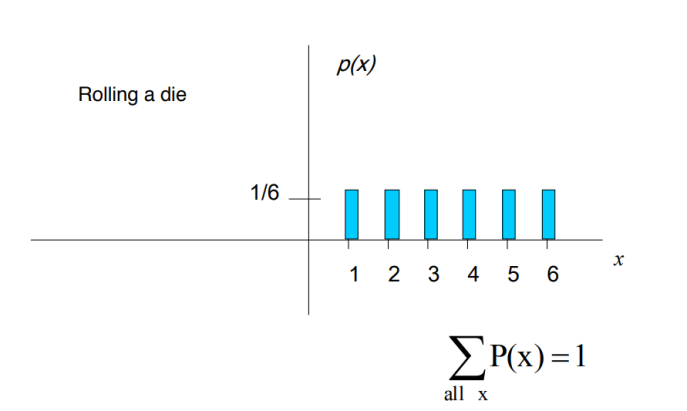
continuous:实数某一区间，统计方法是测出来的/电流电压这种



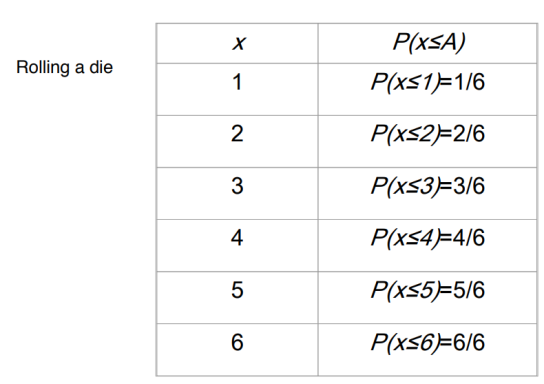
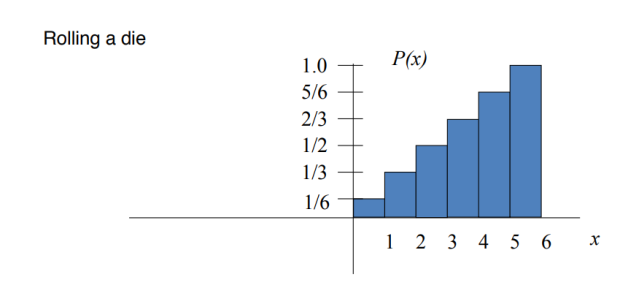
Probability Distributions概率分布

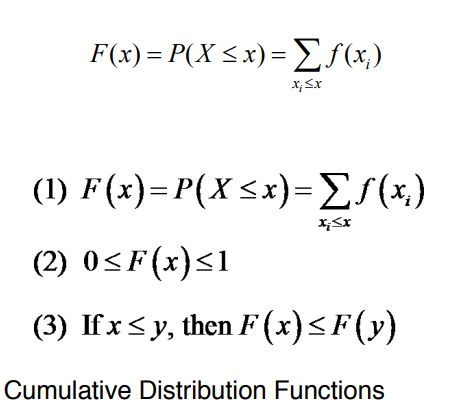
把每个值的X都表示出来

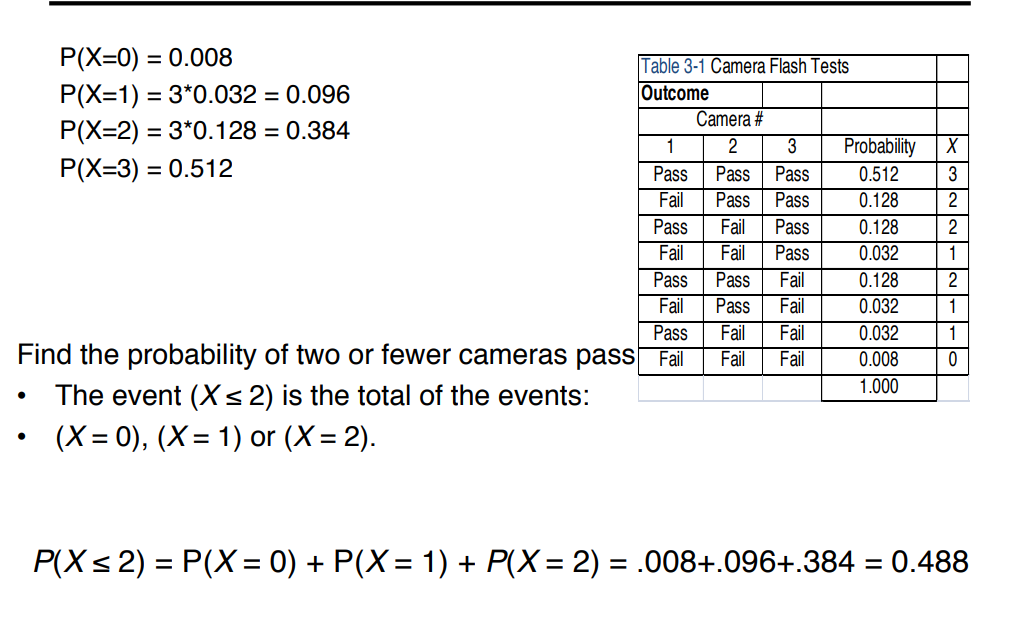
probability mass function:pmf 是离散随机变量在各特定取值上的概率

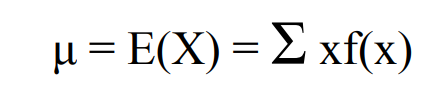
 

Cumulative Distribution Functions：cdf 累积分布函数



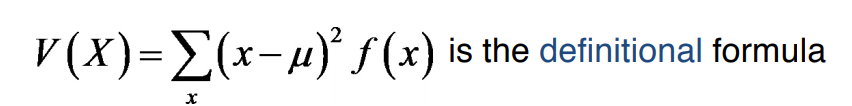


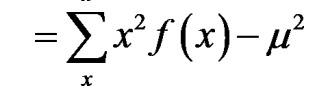


expected value(mean number)：数学期望 小x×pmf（probability）

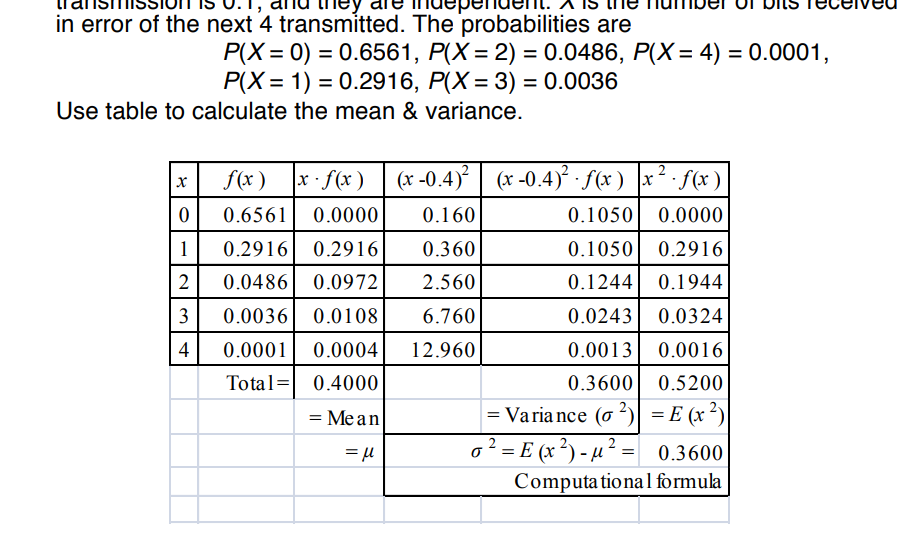
扔一个色子，数学期望是3.5（1+2+3+4+5+6）/6

Variance Formula 方差公式

定义式

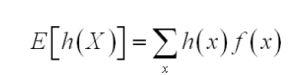
也可以看成（分开来计算方便时用这个） 刻画了离散的程度

就是小x减数学期望的平方×发生概率

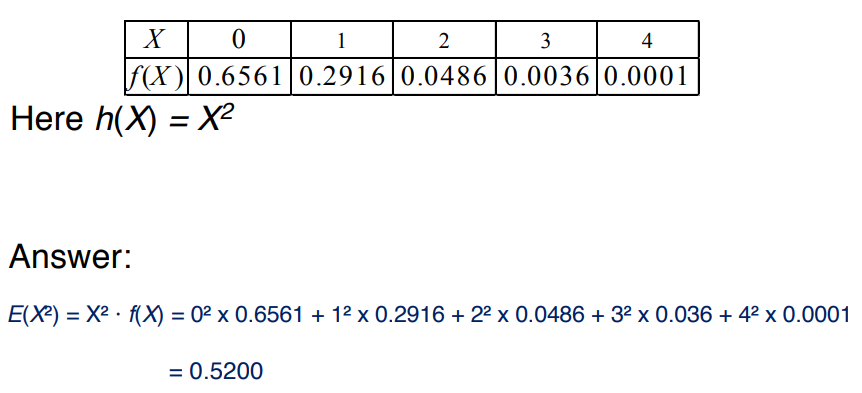


Expected Value of a Function of a Discrete Random Variable：

离散RANDOM VARIABLE的fufnction的数学期望

原来都是单独大X，现在要加个function

如果，那么他的数学期望就是X的variance方差

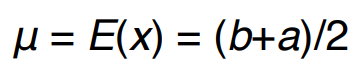


Discrete uniform distribution 离散均匀分布

不同的x1,x2,xn值，有相同的可能性，那么这时候就有discrete uniform distribution,式子为f(x)公式是

如果大写X几率相等，小写x分别为，a a+1 a+2……b（a<b） 例如2 3 4 5 6



数学期望是

方差是

Binomial Distribution 二项分布

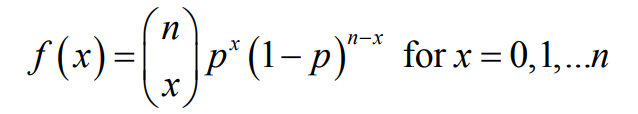
p小写:在一轮中获胜的概率

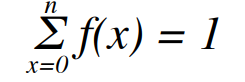
进行n轮

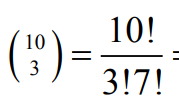
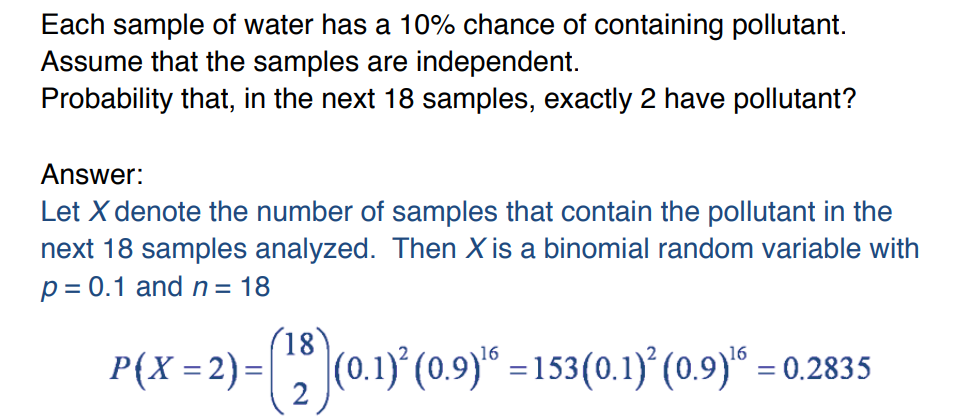
X:random variable，总共获胜的次数

每场获胜互相独立

这里的f(x)指的是总共获胜N次的概率

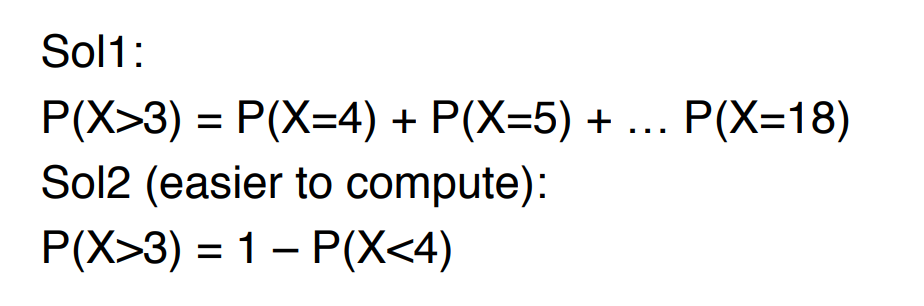


从0轮的f(x)一直加到n轮肯定=1

总结一下就是那个吊东西×成功的次数×不成功的次数

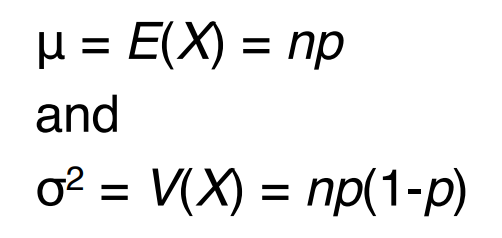
前面的吊东西就是顺序不重要



巧算方法

Binominal mean and variance

如果X是个binominal二项式的random variable， 有着参数p n

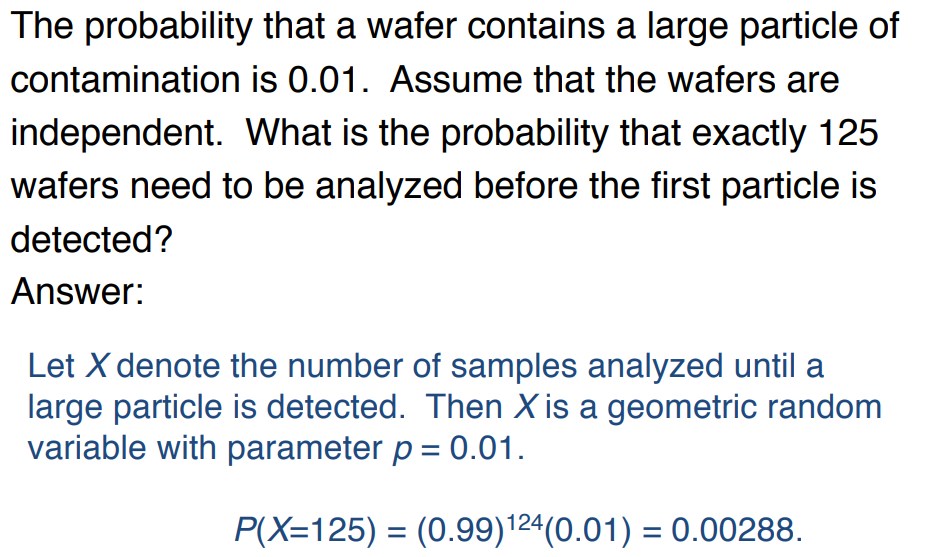
这里X是所有轮结束完以后胜利次数 n是几轮，p是一轮获胜概率

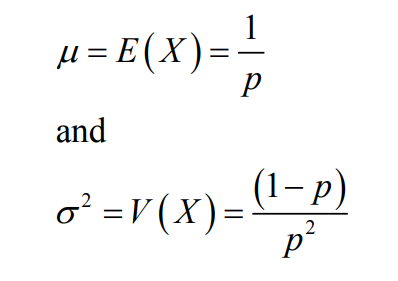
Geometric distribution几何分布

X：随机进行几轮

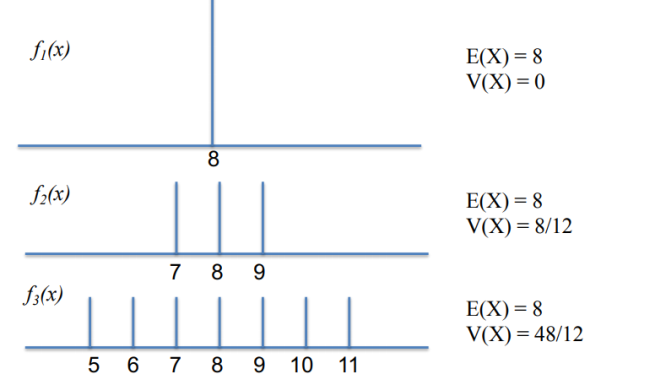
进行几轮获得第一次success

可能性x是成功之前一共经历过的failure次数



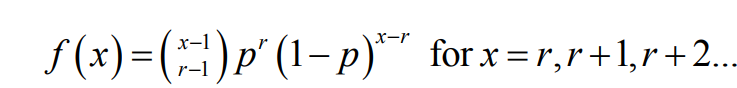


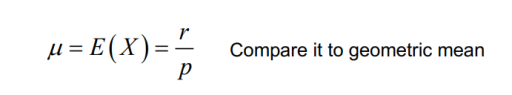
不同的离散均匀分布，即使期望相同，方差不同



negative binominal distribution

negataive binominal random variable X:小x:在r次胜利之前所需要的轮数





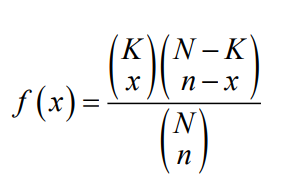
Hypergeometric Distribution.

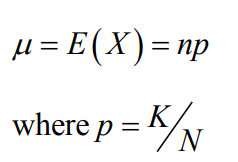
N个object,K个成功，N-K个失败

不重复的从N个object里挑选n个object

random variable X:n中成功数量

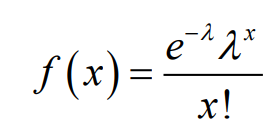
X就是hypergeometric random variable

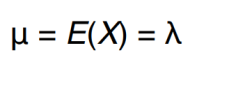
几率是.



Poisson Distribution泊松分布

X是poisson process中event发生次数

几率是

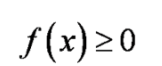
就是poisson process的一个参数

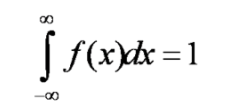
以上都是离散的

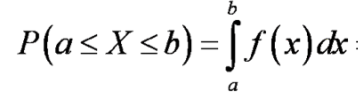
———————————————————————————————————————

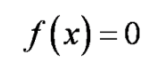
对于continuous random variable X,

PDF (probability density function)

这个function恒大于等于0

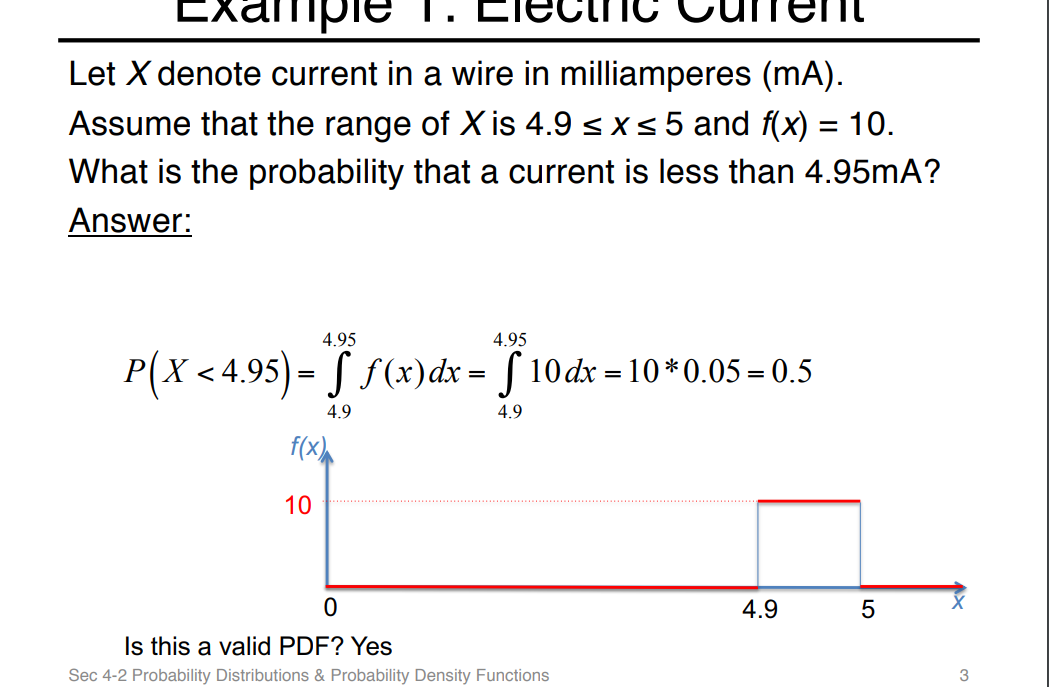
负无穷到正无穷总和等于1

a到b的function的面积

x这个点没area

和离散重要的差别在于fx可以大于1，只要最后总面积等于1就行（因为离散f(x)就是几率，而这里的f(x)×范围才是几率）

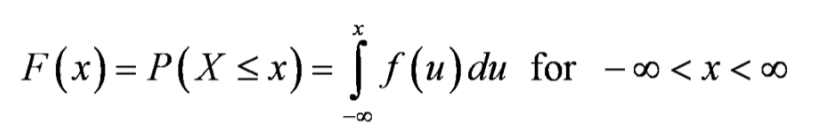
总面积即几率



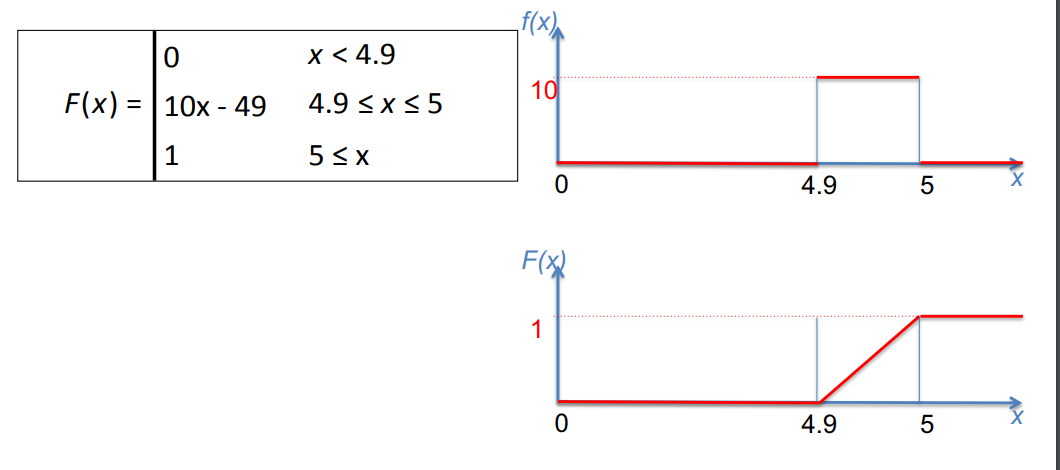
可以看到是不是个valid pdf主要看P是不是在0到1之间

CDF cumulative distribution function

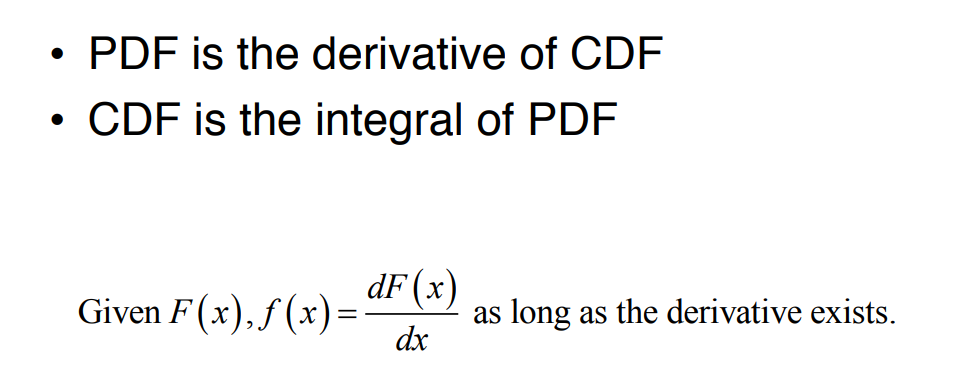
对于continuous variable X,

F(x) 

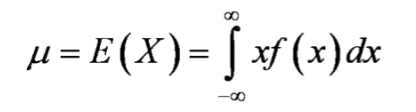
F(x)在不同区间的function公式

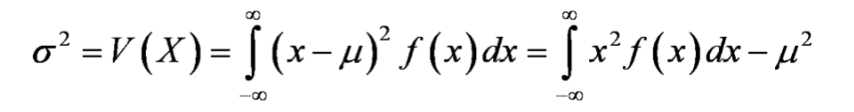


PDF是CDF的导数，CDF是PDF的积分

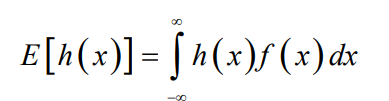


MEAN/EXPECTED VALUE



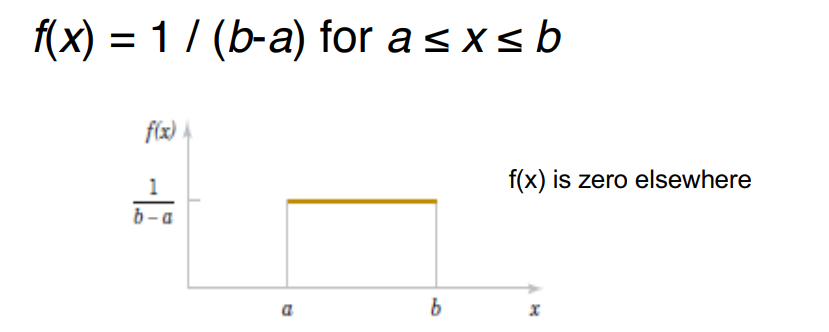
VARIANCE

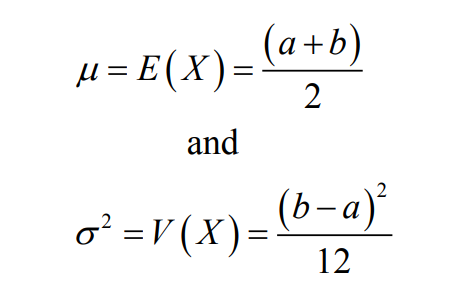




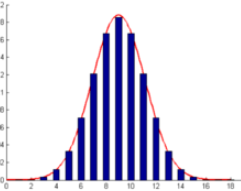
Continuous Uniform Distribution连续均匀分布

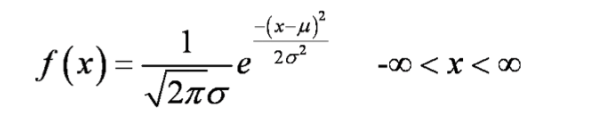
最简单的一种continuous distribution



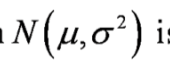


Normal Distribution正态分布



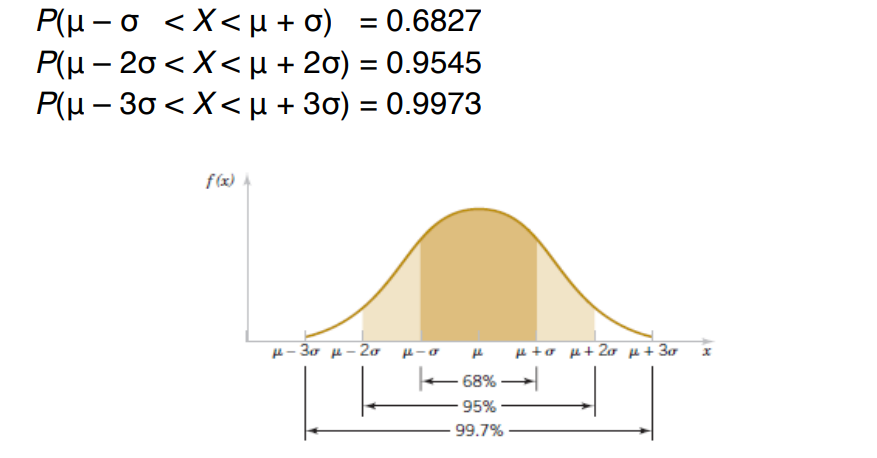
Pdf:



用来表示这种distribution

Empirical law

对于任意random variable，以下式子成立

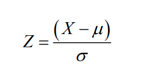


W6 1 ch04

STD

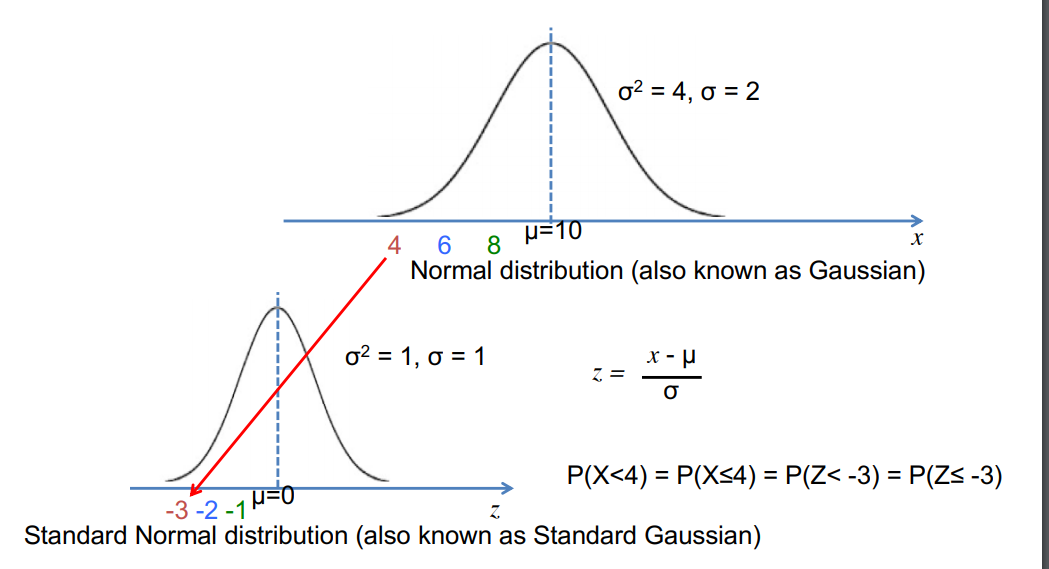
Normal random variable正态随机变量

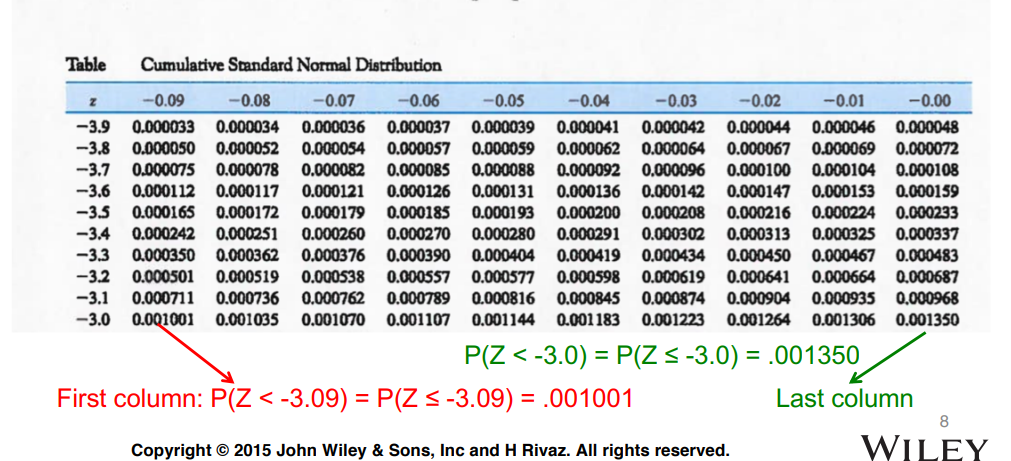
假设X是一个有着mean 为，variance是的正态随机变量（Normal distribution）

而正态随机变量Z（第二个随机变量） ，

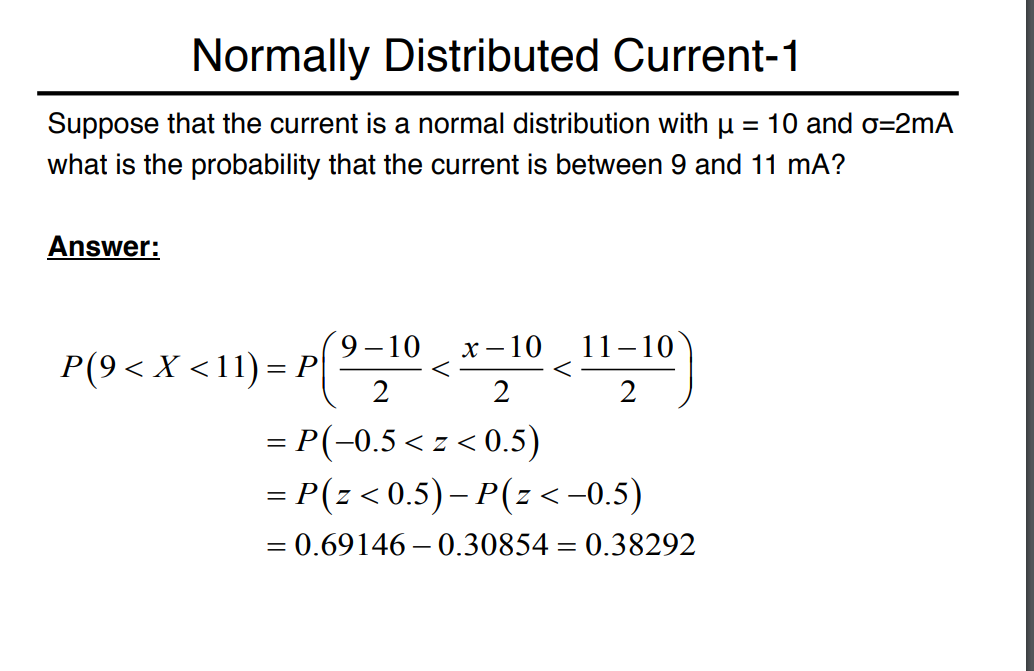
通过使用，我们就能在表格里得到对应的可能性，(standard normal distribution)

这个是两者之间的转换，上图是X，μ=10，方差=4，对应下来的平均几率是x=4的平均几率-10/2

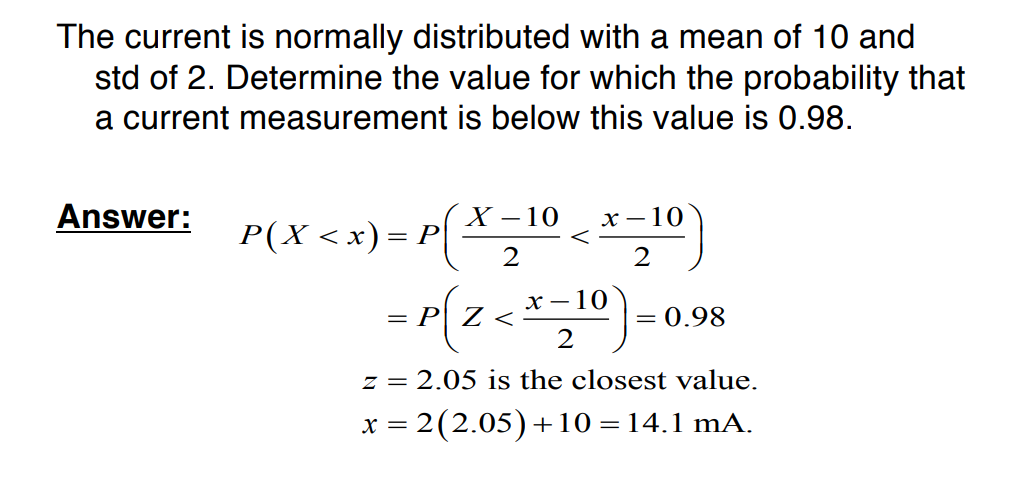




Z=3.09先对应到三，再确认到09



有9到11的X，我们把X转化成Z，所以要-μ/2，注意这里不在乎单位。对应的P还是一样的，然后用0.5和-0.5通过表格得到几率大小

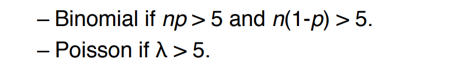


第一步：我们知道几率，用表格找到最近的Z

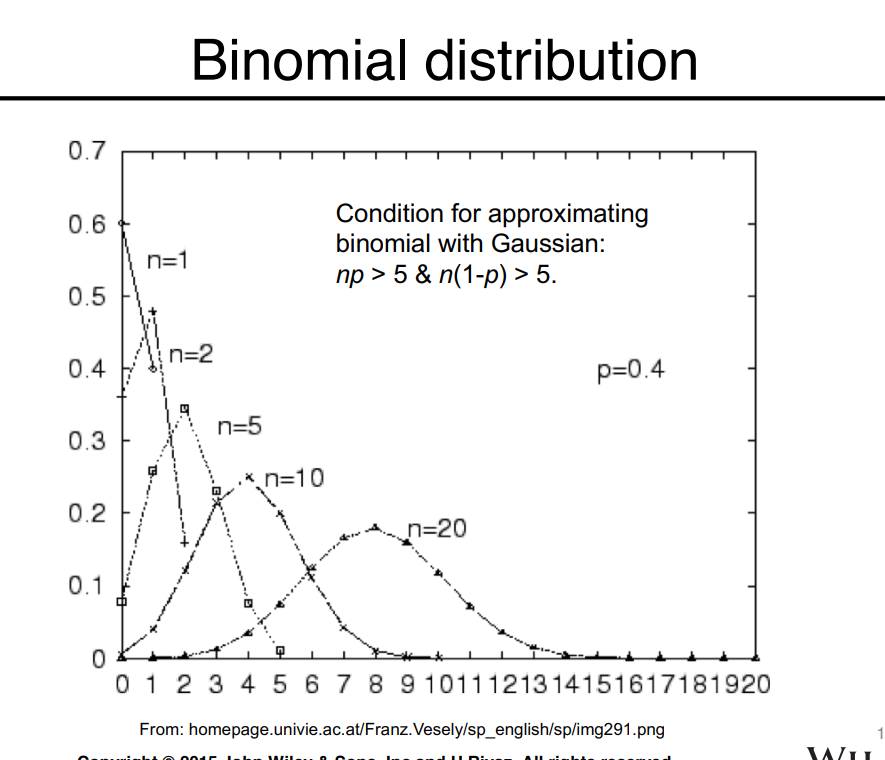
第二步，利用z=x-μ/std得到x

Binomial和possion distribution变得更矮扁和对称，随着他们的mean value不停上升

对于手工计算，normal approxiamation是非常实用的，当

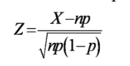


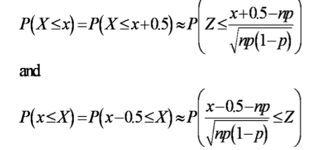
否则就要用计算器计算



可以看到，随着n越来越大，越来越对称，越来越接近标准形状bellshaped

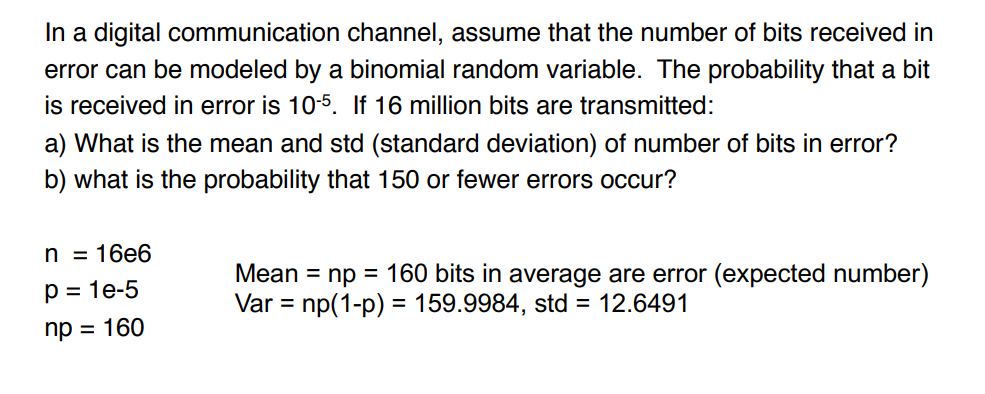
如果X是binomial random variable有着参数n和p

那么就近似于standard normal random variable，为了近似的得到二项可能性通过normal distribution，我们需要应用continuity correction

这个近似法Z，只有当np与n1-p都大于5时才有效

注意这个是必须使用的，X<要+0.5,大于要-0.5

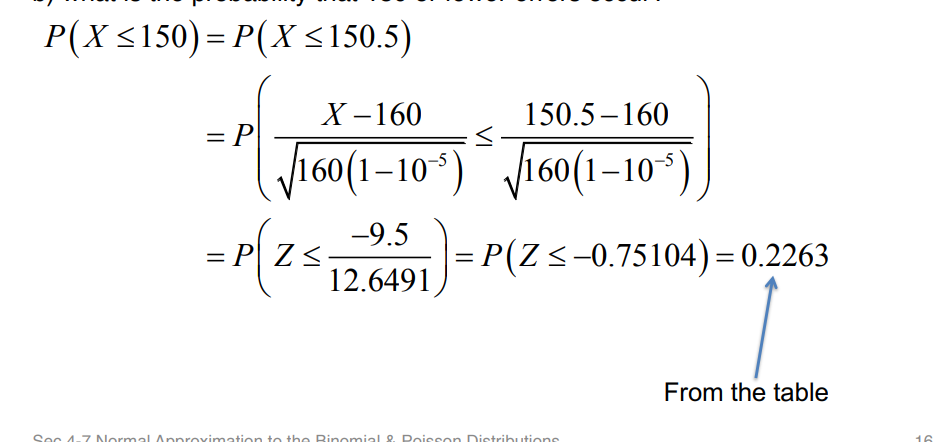
这里的n指几轮，p指一轮获胜概率



Np皆有，算np得到mean,np1-p得到var，

因为是

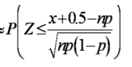
小于，所以必须加0.5



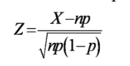
注意这个Z<=多少是对应的表格

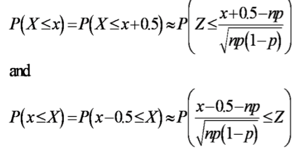
例子

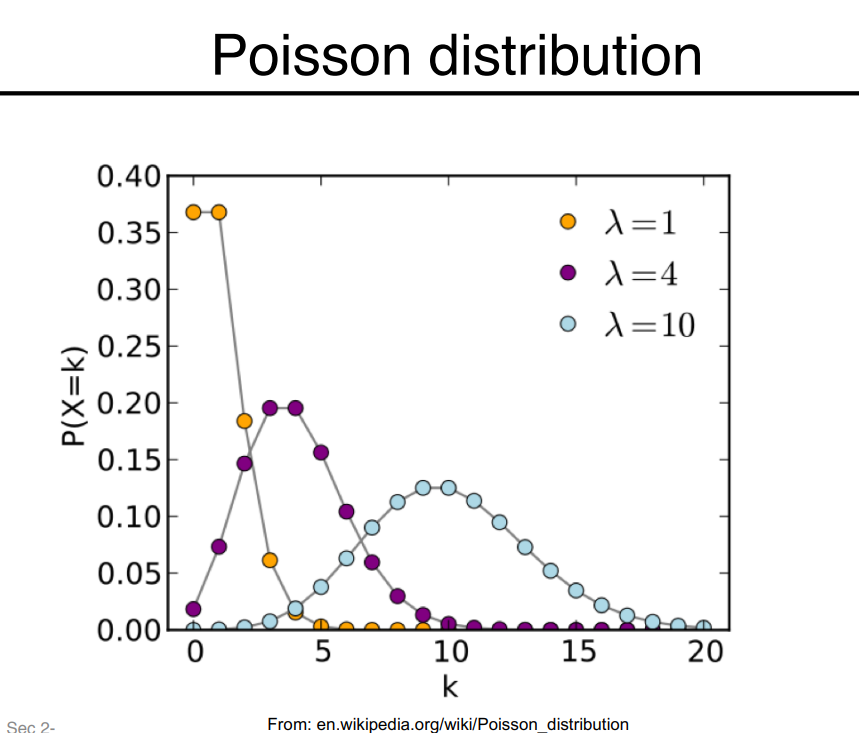
反正就是让你求x<=150，然后你可以突然加上0.5，然后代入式子，然后就可以查表了，第二行括号其实就是，X<=150.5 = P<=160-150.5到160之间的P，而左边的就是我们所要求的Z

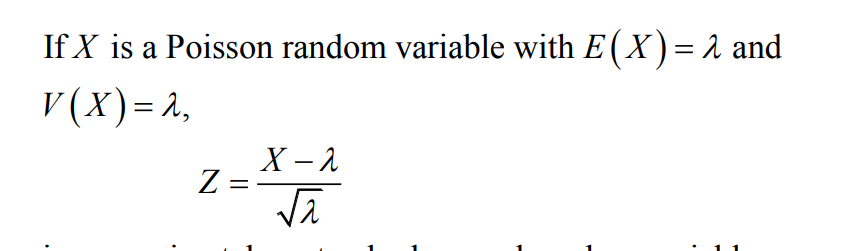
其实可以直接代入式子，x=150,右边的式子就是

总结：

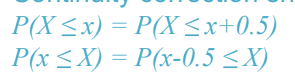
Binomial求P,把X转化成Z要遵循,然后小于+0.5，大于-0.5

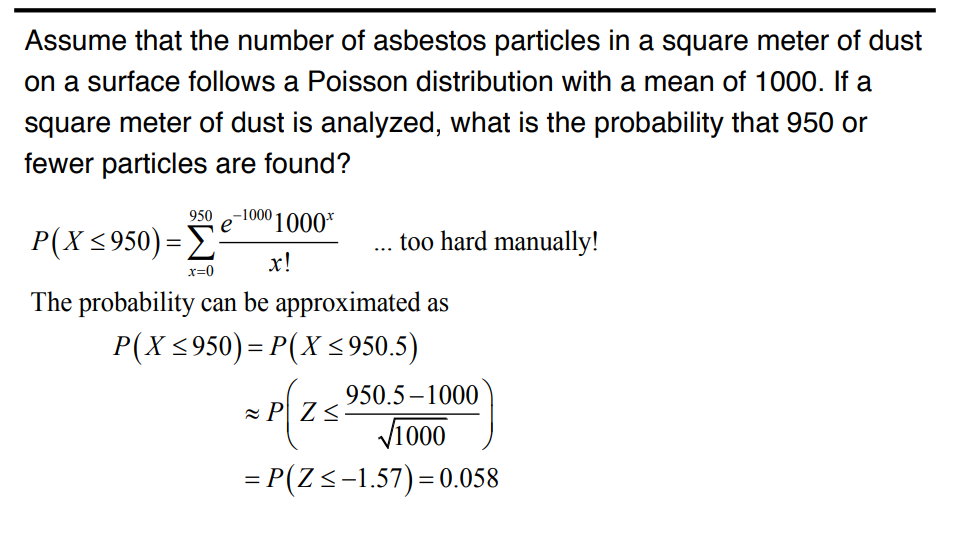




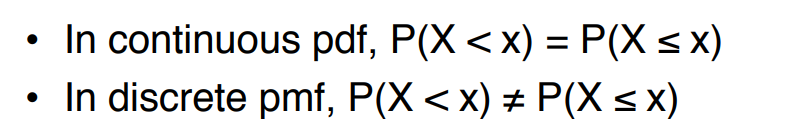


只有在时生效

这个定律同样能使用



Poisson确定，小于确定，950.5代入式子，OJBK

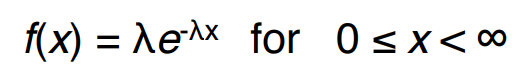


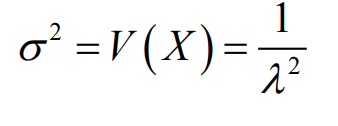
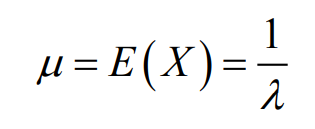
对于连续pdf，与对于离散pdf

所以说对于离散型approxiamate的时候，我们要注意是还是discrete distribution to standard normal distribution

而对于连续型来说，不用注意，normal distribution to standard normal distribution

Exponential distribution definition指数分布的定义

可能性



注意：对expotional dixtribution,mean =std

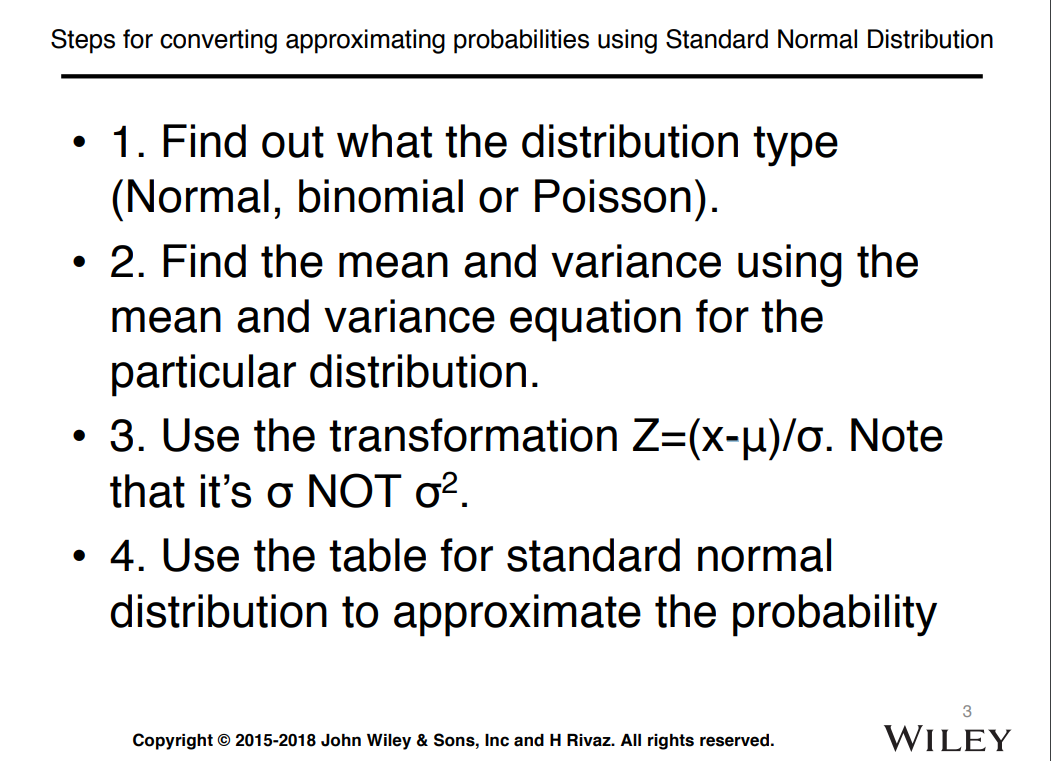
对poisson，mean=variance

第一步，查找出那种distribution(normal,binomial,poisson)

第二步，用公式得到mean 和variance

第三步，用Z的变形式，

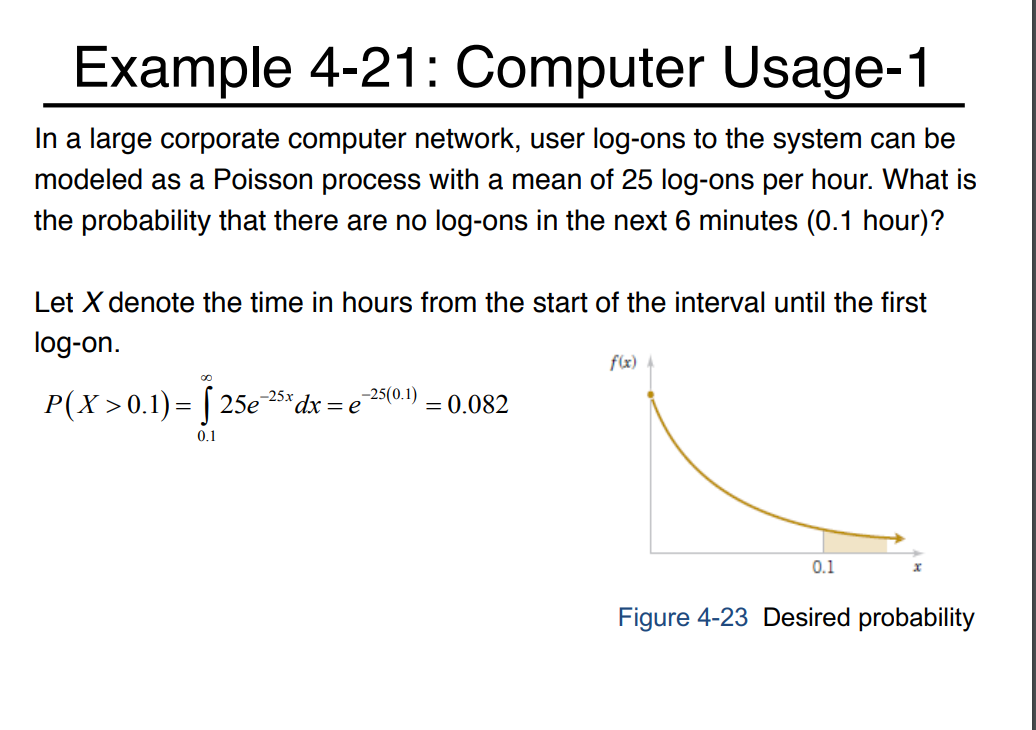
第四步，用表



Poisson分布和expotional 分布的区别

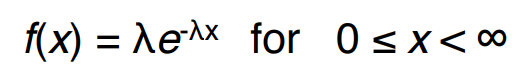
http://www.woshipm.com/pmd/163461.html

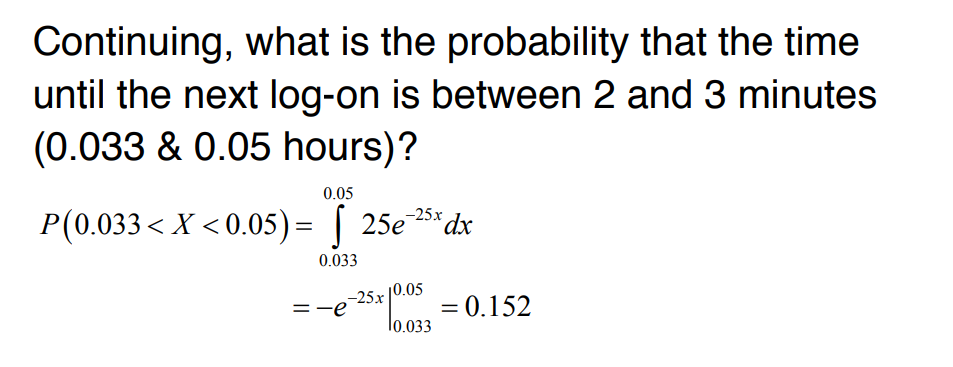
例子



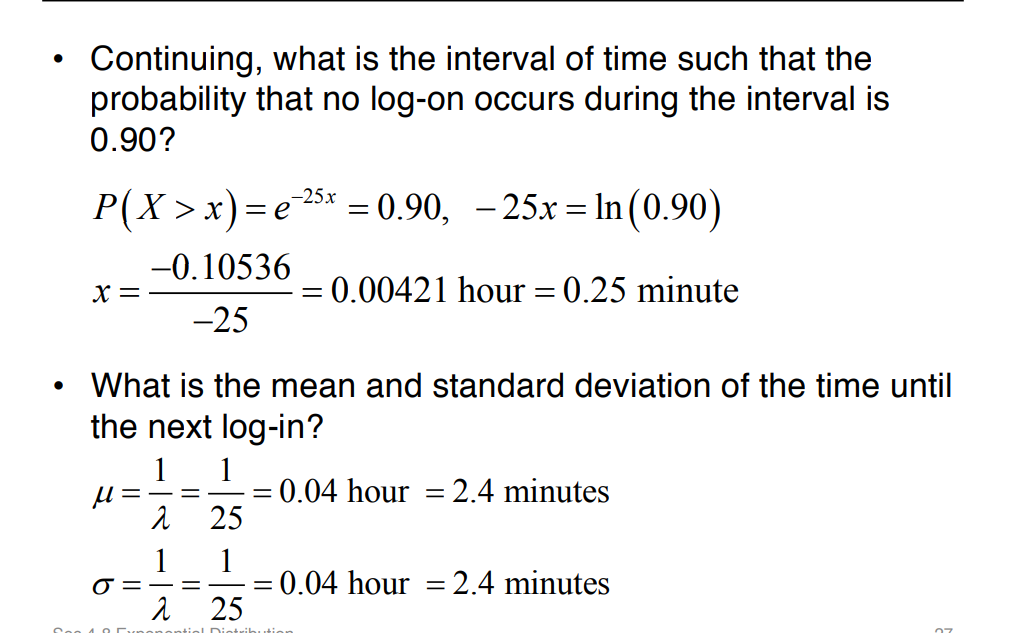
已知事件一小时发生25次，那么6分钟一次也没发生的几率

在0.1以后发生第一次的几率

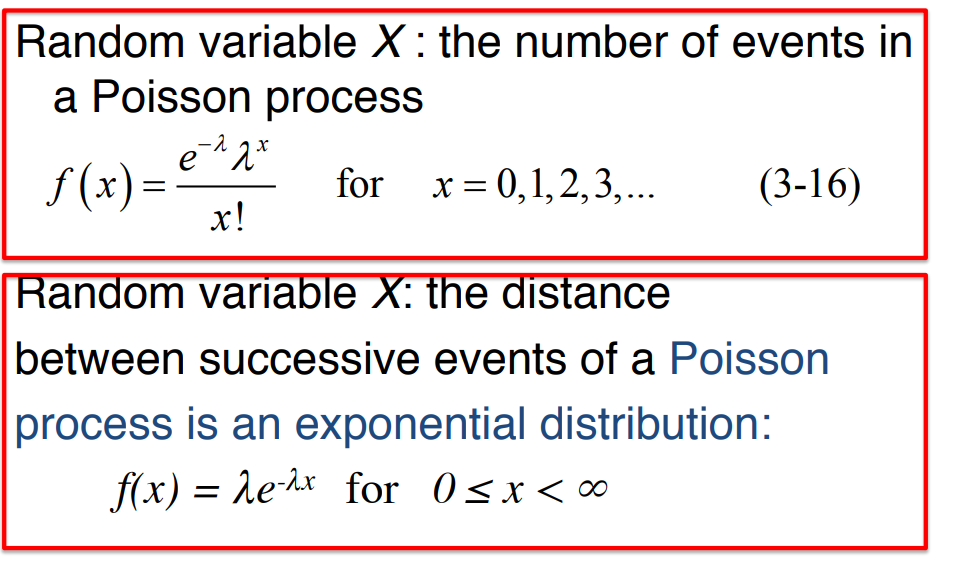
要加个积分



反正要求什么时候发生，就求从多少多少分钟到多少分钟，一定要用时间为积分



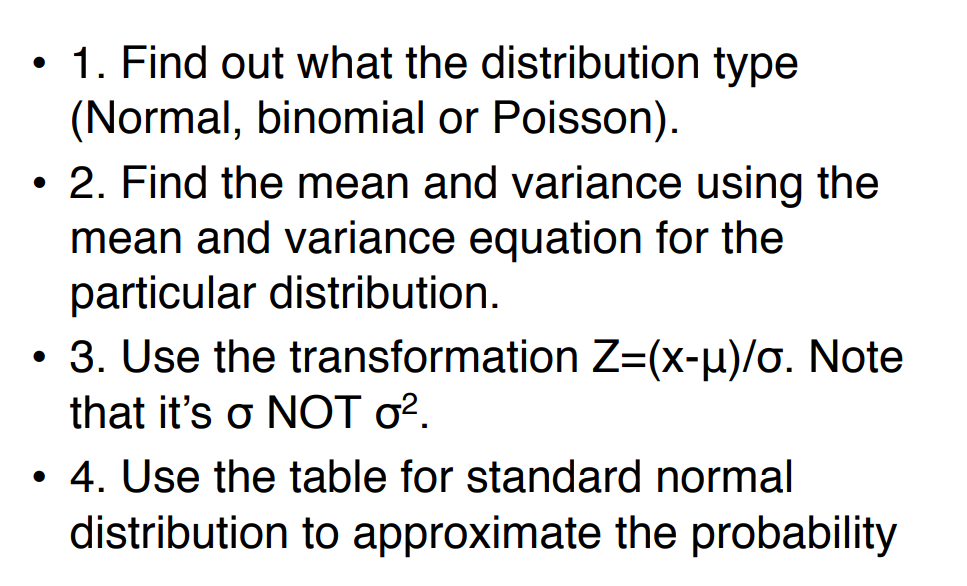
其实和第一题差不多



对于poisson来说，X指的是一个poisson过程里发生几次

对于exponential来说，X指的是poisson过程中第一个成功离出发时间的距离

几部曲



1. 找到distribution的type
2. .找到对应的mean 和variance
3. 利用对应的XZ转化共识，binomial或者poisson要+-0.5
4. 查表