Borel域{}，由此可以得到左闭右开，闭区间，单点集，左开右闭，开区间（一.64）

古典概率模型与计算：

**一.古典概型**：有限性；等可能性。(加法原理、乘法原理)

|  |  |  |
| --- | --- | --- |
| N个不同元素取r个的不同结果数 | 不可重复 | 可重复(放回) |
| 计次序 |  |  |
| 不计次序 |  |  |

可重复不计次序可视为满足方程x1+x2…xn=r的非负整数解，或的展开式中不同项数，或r个不可分辨的球分成n组（可空）的方法数，或加上n个球，这样每组不空

概率公式

1. 二项分布：放回抽样，N个东西中有M个B，取出n件中有m件B的可能，
2. 超几何分布：不放回抽样，N个东西中有M个B，取出n件中有m件B的可能，

**二.几何概型：**有限性；等可能性（度量相等时）

**三.概率的公理化定义与性质**

设为样本空间，F为事件域，任意事件A属于F，满足：①非负性②正则性③(互不相容事件)可列可加性④单调性⑤差事件⑥加法公式

加法公式推广：-

对递增事件有下连续性，递减事件有上连续性；(一.143)

概率空间

**四．条件概率**

公式，B发生条件下，A发生的概率，具有非负性、正则性、（互不相容事件）可列可加性

乘法公式：;

Polya模型(一.167)

全概率公式：

贝叶斯公式（逆概率问题）：(机器学习)

独立性：或，互不相容与独立相矛盾（推广：）

一组事件相互独立，则两两独立，反之不一定

两个试验间任意结果相互独立，则两个试验独立，类似有n个试验独立，当每个试验相同，为n重独立重复实验，如果结果只有两个，则称为n重Bernoulli试验

随机变量

**分布函数**，记为X~F(X)

具有单调递增性、有界性、右连续性（）

;;;

离散型随机变量—分布列，分布函数为跳跃（阶梯）函数，间断点右连续

连续型随机变量—概率密度函数，

可以表示为变上限积分函数的分布函数为绝对连续函数，但分布函数连续不一定意味着RV是连续型

某点处概率为0，概率密度函数不唯一

若F（x）的导数在除有限或可列个点外存在且连续，则X为连续型RV（在某一点的概率不可能大于零）

非离散型分为：连续型、奇异型、其它（混合型..）

**二项分布：在n重伯努利试验中，事件A发生的次数**

概率

**Poisson分布：一本书中的印刷错误；某天进入商场客户数**

**独立增量性（不相交时段内，发生数相互独立）；时齐性（在长度相同时段内，发生数统计规律相同）；普通性（在极短时间内，发生数概率近似与时长成正比）**

分布列

时，二项式分布趋近于poisson分布

**负二项分布（巴斯卡分布）：伯努利试验第r次成功时实验次数X**

概率分布

**几何分布：伯努利试验首次发生时的试验次数X**

概率分布

具有无记忆性

**超几何分布：N个中有M个特殊，取n个中有X个特殊**

(以下为连续型分布)

**指数分布**

概率密度函数

分布函数

具有无记忆性

Poisson分布与指数分布的关系(二.97)

在任意时长t内发生故障次数满足poisson分布P()，则相邻两次故障时间间隔满足指数分布

**正态分布：标准正态X~N（0,1）**

一般正态密度函数,即为X~N（,），越小，图形越尖锐陡峭

一般正态标准化：~N（0,1），3原则

**Gamma分布：**

密度函数,,,

**数学期望：**

对于离散型RV，需要级数绝对收敛，即

对于连续型RV，需要积分绝对收敛，即

则称数学期望存在

**方差：Var(X)=D(X)=E[(X-EX)^2]=E(X^2)-E(X)^2（只需判断E(X^2)是否存在即可），标准差：(X)=**

**D(aX+b)=D(X)**

**二项分布数学期望与方差：np，np(1-p)**

**几何分布数学期望与方差：，**

**Poisson分布数学期望与方差：，**

(以下为连续型)

**均匀分布数学期望与方差：，**

**指数分布数学期望与方差：，**

**正态分布数学期望与方差：，**

**柯西分布数学期望不存在**

**Lotus:E(Y)=E(g(x))=…**

**E(g(X)+f(X))=E(f(X))+E(g(x)),E(aX+b)=aE(X)+b**

**当a=E(X)时，E[(X-a)^2]最小**

**切尔雪夫不等式：（期望为，方差为）**

**K阶矩:；变异系数：**

**分位数：，为此分布的下侧p分位数**

**中位数：，m为中位数**

**偏度系数：；峰度系数：**

**设Y=g(X)，求其概率密度函数：①直接法：先求其分布函数**

1. **变换法：**

**多维随机变量**

**联合分布函数：单调性，有界性，右连续性，任意矩形非负性**

**多项分布（二项式推广）：n次独立试验，每次试验有r种可能结果，A\_i为其中一种结果发生的次数，（n为和）**

**多维超几何分布：N个球，N\_i个i号球，取n个球，其中X\_i为其抽出个数，**

**多维均匀分布**

**二元正态分布：**

**有五个参数**

**边际分布:, 对于正态分布，不同的有相同的边际分布**

**独立：，对于离散型；对于连续型等价于几乎处处成立（必须本质定义于正矩形区域）（或者看是否可分离变量）**

**对于正态分布，独立的充要条件是**

**若X与Y相互独立，则g(X)和h(Y)相互独立**

**多维离散型RV的卷积：**

**X，Y相互独立，且为离散型，设，，那么Z=X+Y的分布列为**

**即为卷积**

**Poisson分布的可加性：设X与Y独立，分别满足Poisson分布、，那么其和的分布列为，n个同样适用**

**二项分布的可加性：设X~B(n,p)，Y~B(m,p)，X与Y独立，则X+Y~B(n+m,p)，即B(n,p)\*B(m,p)=B(n+m,p)**

**几何分布的卷积：设X~Ge(p)，Y~Ge(p)，则X+Y满足负二项分布Negbin(2,p)，，有**

**最大值与最小值的分布：**

**对于最大值分布的一般情况，有，与相互独立，则，对于n个也相仿**

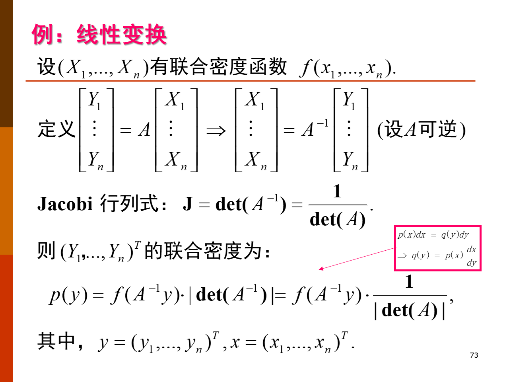
**对于最小值分布的一般情况，有，与相互独立，则**

**当二者不独立时，对于n个也相仿**

**变换公式：**

**设，欲求其联合密度函数，已知其逆函数为**

**那么有，更高维同理，**

****

**RV和的分布：①增补变量法②分布函数法（利用联合分布函数）**

**（当与相互独立，则的密度为）**

**正态分布的可加性：**

**若，且相互独立，那么，，但是要注意（因为相关）**

**均匀分布不具有可加性**

**指数分布不具有可加性**

**Gamma分布：，称为Gamma分布，形状参数，r为尺度参数**

**RV之差的分布：**

**多维RV的特征数：**

**设，**

**2D LOTUS公式**

**E(Z)=E[g(X,Y)]=（离散），（连续）**

**E(X+Y)=E(X)+E(Y)（X与Y不需要相互独立）**

**Max{X,Y}=，Min{X,Y}=**

**相互独立RV之积的期望：**

**设X与Y相互独立，则E(XY)=E(X)E(Y)**

**相互独立RV之和的方差：D(X+Y)=D(X)+D(Y)**

**协方差：**

**对于不相互独立的RV的和的方差，有**

**E[(X-EX)(Y-EY)]可以用来衡量两个量之间的相互关系**

**定义协方差为Cov(X,Y)=E[(X-EX)(Y-EY)]**

**特别的，有Cov(X,X)=D(X)**

**正相关：Cov(X,Y)>0，X与Y同增减**

**负相关：Cov(X,Y)<0，X与Y增减性相反**

**不相关：Cov(X,Y)=0，二者不线性相关**

**协方差的计算：Cov(X,Y)=E(XY)-E(X)E(Y)**

**定理：若X与Y相互独立，则X与Y不相关，即Cov(X,Y)=0，反之不一定成立（不相关不能推出独立）**

**（例子：一般，若X服从对称分布，或|X|，则X与Y不相关，但不独立）**

**不相关是就线性关系来说的，独立性是就一般关系来说的**

**协方差（RV和）的方差推广：**

**协方差性质：**

1. **Cov(X,Y)=Cov(Y,X)**
2. **Cov(X,a)=0**
3. **Cov(aX,bY)=abCov(X,Y)**
4. **Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)**

**双线性性：**

**相关系数：定义（X,Y）的相关系数为**

**相关系数是规范化RV的协方差**

**对于满足二维正态分布的（X,Y）~其相关系数为**

**定理：对于二维联合正态分布的两个RV，其不相关与独立等价，都是**

**Cauchy-Schwartz不等式：**

**这样有定理：**

**相关系数性质：；的充要条件为X与Y几乎处处有线性关系**

**若要找aX+b来拟合Y，那么有，相关系数可以拟合线性近似程度的好坏，最佳线性预测为**

**随机向量的数学期望与协方差阵：设，存在，定义X的数学期望为，其协方差阵为Cov(X)=C，其中，且其对角线元素，可以表示为，协方差阵是一个对称的非负定矩阵**

**多维正态分布：（未整理）**

**条件分布函数：**

**连续RV分布的条件密度函数为**

**由此也可以利用乘法公式，求解联合密度**

**全概率公式的密度函数形式可以由一个边际求另一个边际**

**连续场合的Bayes公式**

**条件数学期望：**

**离散：**

**连续：**

**若相互独立，那么**

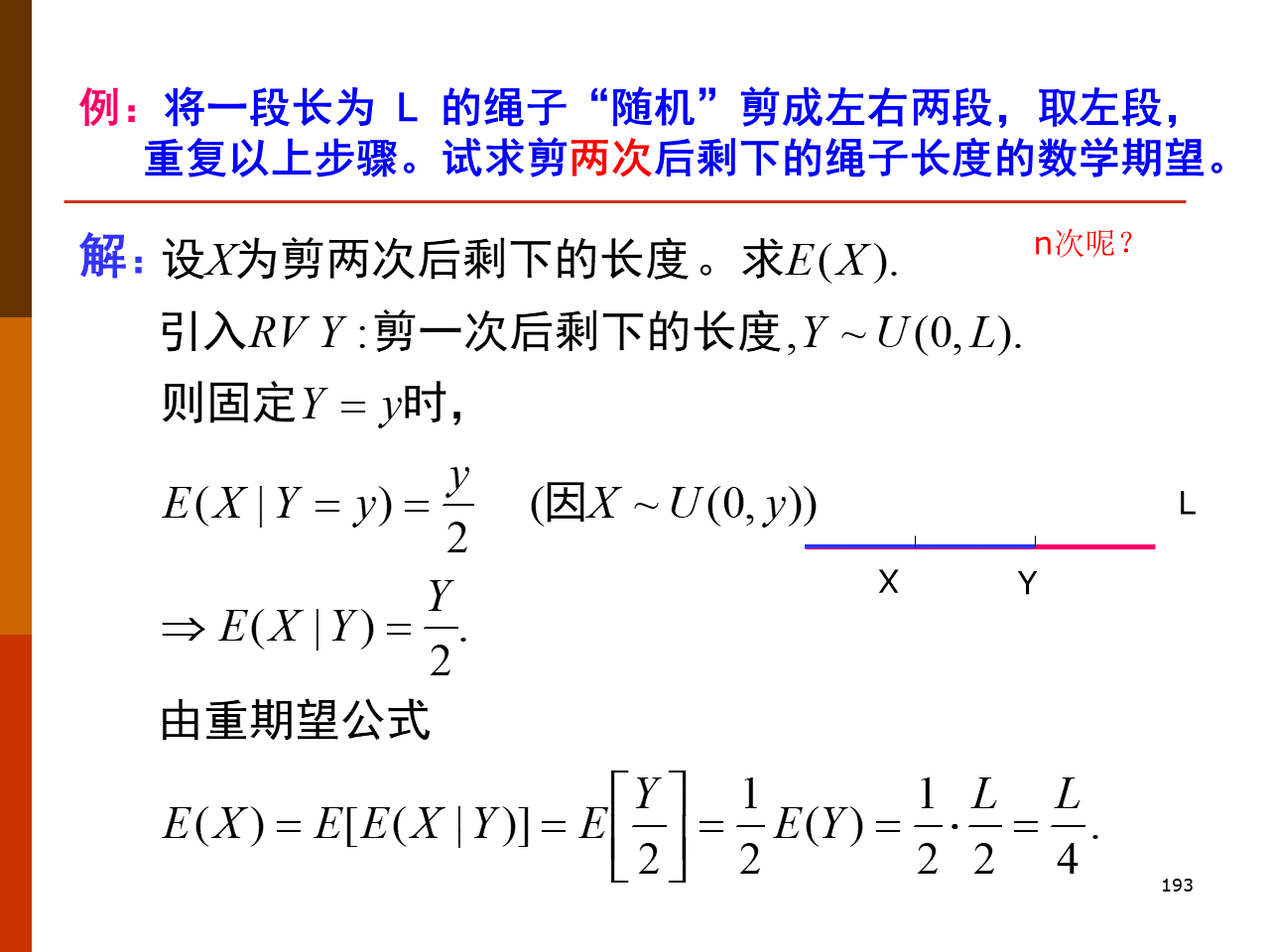
**作为RV的条件期望：条件期望是关于y的函数，记**

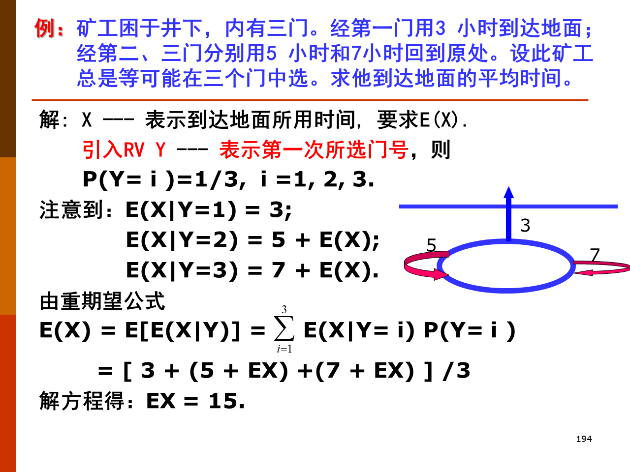
**，那么定义新的RV为**

**重期望公式：，J为某个组分，后面J为权重，相当于概率**

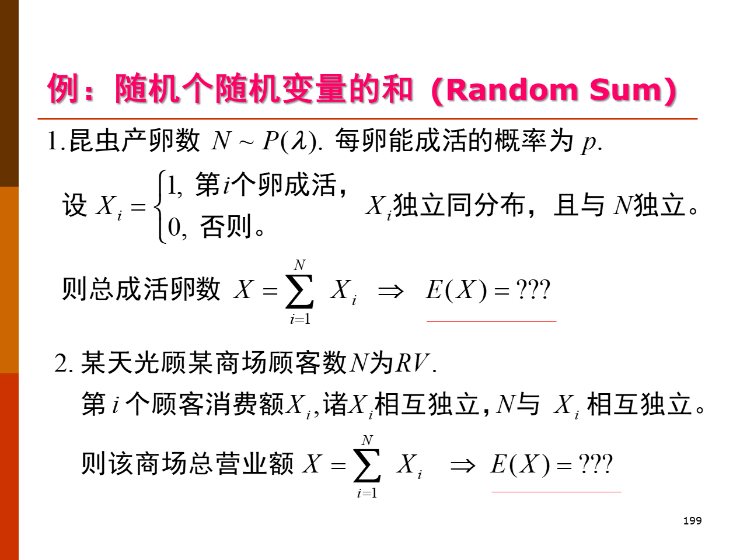
**设为二维RV，且E(X)存在，那么**

**所以利用重期望公式，可以引入RV Y，利用重期望公式，求解**

****

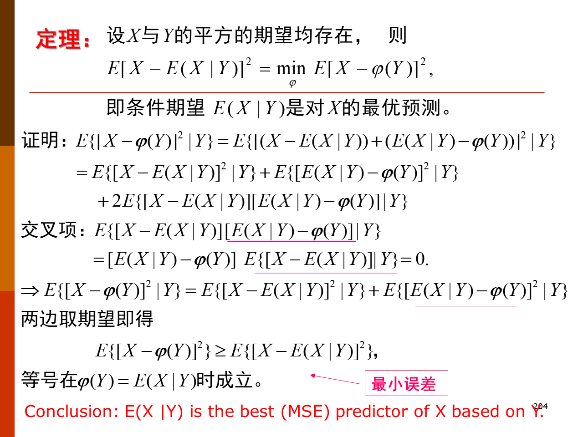
****

**可以用来求解随机个随机变量的问题：**

****

**若相互独立同分布，那么有**

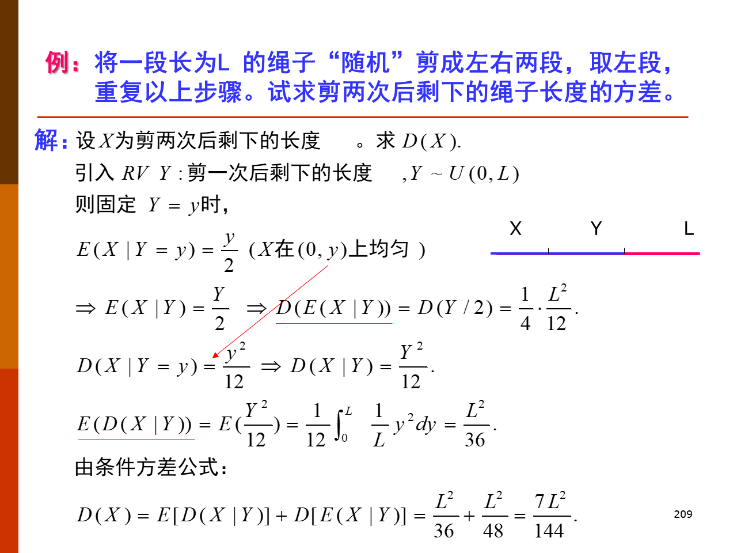
**定理：设X与Y的平方的期望均存在，则**

****

**条件方差：**

**那么有**

**随机个随机变量和的方差：**

****

**第四章 大数定理与中心极限定理**

**特征函数不做要求**

**母函数：对于，有母函数，为X的母函数**

**复随机变量：，**

**若与相互独立，那么与相互独立**

**特征函数的定义：**

**二项分布的特征函数：**

**Poisson分布的特征函数：**

**标准正态分布的特征函数：**

**特征函数性质：**

1. **，**
2. **共轭对称：**
3. **若，则**
4. **特征函数在上一致连续**
5. **半正定性：对于任意个复数，有**
6. **设与相互独立，那么，这样分布函数的卷积变为特征函数的乘积**
7. **设存在，那么的特征函数可以微分n次，且，特别的有**

**逆转公式：设分布函数的特征函数为，与是的连续点，那么**

**另外**

**唯一性定理：分布函数由其特征函数唯一确定**

**若特征函数的模可积，那么相应的分布函数导数存在且连续**

**RV的两种收敛：①依概率收敛（用于大数定律）②依分布收敛（用于中心极限定理）**

**对于n重伯努利试验，设事件A发生次数为是RV，那么考虑频率，有，，均值不变，而方差趋于0**

**在大量实验中，随机事件表现出频率的稳定性**

**这样，有大数定律：{}为频率序列，p为事件A发生的频率，那么，有，或者，那么称为{}依概率收敛于p，记为**

**{}为频率序列，p为事件发生的概率，若，则称{}以概率1收敛于p，记为**

**{}为频率序列，考虑标准化的随机变量，那么的分布函数的极限行为是，其分布函数是正态分布**

**定理：设，，函数在连续，则**

**弱收敛、按分布收敛：**

**定义：设RV X的分布函数为，的分布函数为，若对的任一连续点x，都有，则称{}弱收敛于，记为，也称为RV序列{}按分布收敛于，记为**

**弱收敛比点点收敛更弱**

**设，则，反过来不一定成立**

**定理：若c为常数，则等价于**

**以概率1收敛强于依概率收敛**

**大数定律：**

**定理（Markov不等式）：设X为取非负值的RV，则对于任何常数，有**

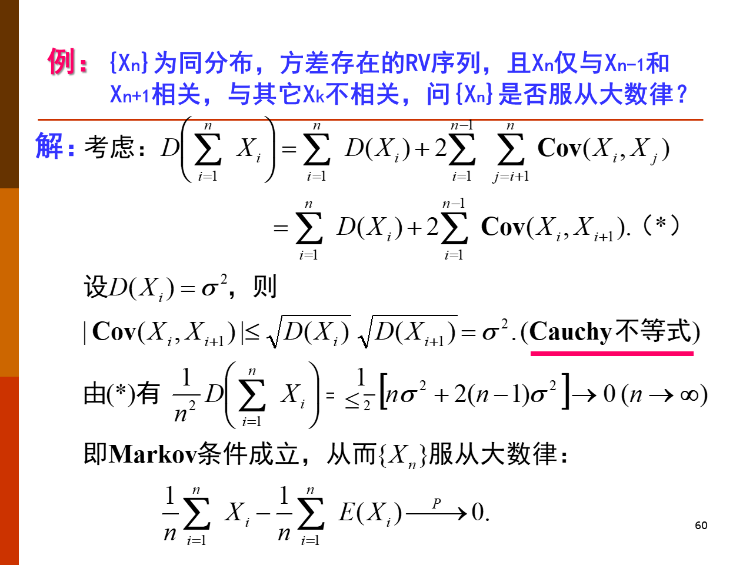
**定理（Chebyshev不等式）：设RV X有数学期望，，那么对于，有**

**定理（Bernoulli大数律）：为n重伯努利试验中事件A发生的频率，p为其发生的概率，则依概率收敛到p**

**大数定律：（序列服从大数定律的定义）若{}为RV序列，称其服从大数律，如果**

**Chebyshev大数律：设{}为相互独立的RV序列，各个方差均存在，且有公共上界，则称该序列满足大数律**

**Markov大数律：设{}为RV序列，和存在，若，则其服从大数律**

****

**辛钦大数律：设RV 独立同分布，有数学期望，则序列{}服从大数律**

**大数律的应用：**

1. **计算面积**
2. **估计积分**

**中心极限定理：**

**大数定律是研究RV的平均的稳定性或者渐进性**

**中心极限定理是研究独立RV的和的极限分布**

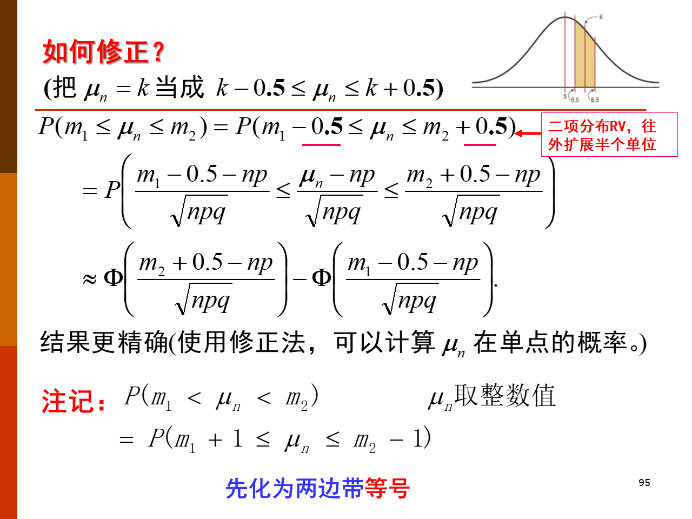
**独立RV的和的精确分布可以利用卷积分布，某些情况下可以利用可加性**

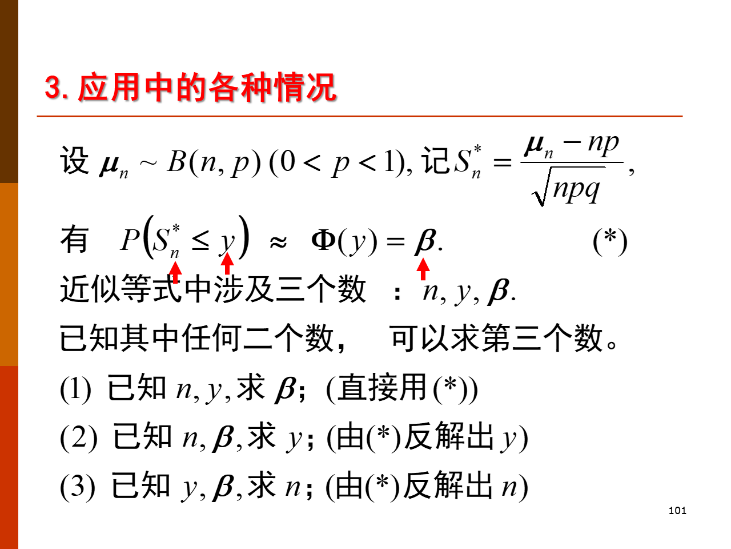
**设{}是独立同分布RV序列，，，记，那么有，**

**De Moivre-Laplace中心极限定理（二项分布的正态逼近）：设RV 为n重伯努利试验中事件A发生次数，那么令，满足标准正态分布**

**中心极限定理比大数定律更加精确，回答了具体的趋近过程**

**中心极限定理的应用：**

1. **二项分布的正态逼近**
2. **连续性修正：二项正态分布作为离散分布二项分布的近似时，应该做修正：**



**独立不同分布下的中心极限定理（了解）**

**第五章 统计量及其分布：**

**总体与样本：**

**总体：研究对象的全体，数量化后对应某个RV X，其分布称为总体分布**

**个体：总体中的每个元素**

**样本：从总体中随机抽取的部分个体的集合，，n为样本容量**

**样本的二重性：**

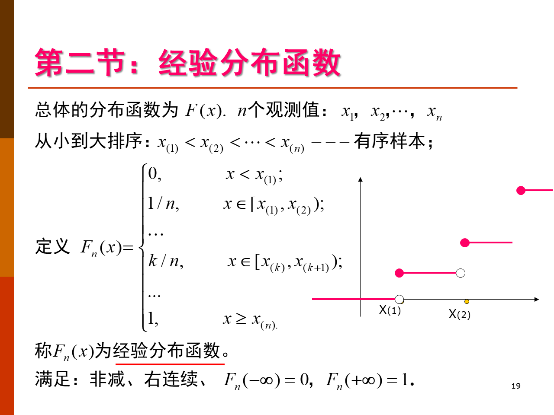
**随机性（抽样前）：无法预知样本具体值，用RV 表示**

**确定性（取样后）：得到n个确定的观测值，用表示**

**简单随机取样（随机性、独立性）**

**样本联合分布函数：**

**设总体具有分布函数，那么样本的联合分布就是随机向量的联合分布，有**

****

**统计量：**

**通过样本，构造适当的函数，对总体进行估计和推断**

**定义：设是来自总体的一组样本，设是样本的函数，且不含未知参数，则称T为统计量，统计量分布称为抽样分布**

**统计量也具有随机性和确定性**

**样本均值：定义为，此时有为样本均值的观测值**

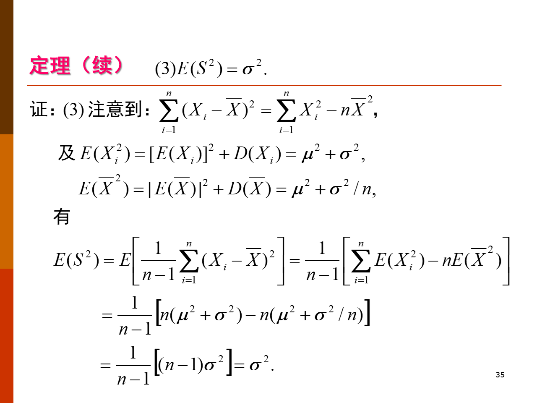
**有的最优解为**

**定理：①若总体分布为，则的精确分布为②对于任意分布，若，，那么n较大时，其近似分布为**

**样本方差：，观测值，样本标准差，观测值相似，定义无偏样本方差：**

**偏差平方和公式：**

**设总体X的均值、方差存在，即，，那么为样本均值，为样本方差，有，，，样本之间相当于独立同分布**

****

**样本矩：定义样本k阶原点矩：，相应的，有观测值，样本k阶中心矩**

**设总体X的k阶矩存在，那么按概率收敛于，连续函数也一样有**

**样本矩：当总体分布不对称时，只用平均值和方差刻画不足，需要引入样本偏度和样本峰度**

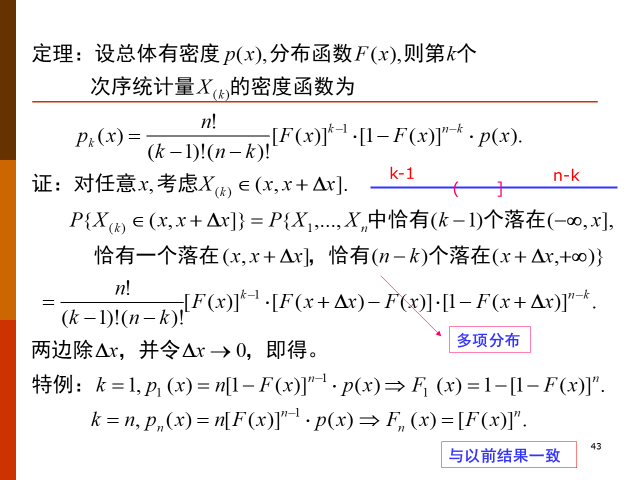
**样本偏度：，反映总体分布的对称**

**样本峰度：，反应总体分布在峰值附近的陡峭程度**

**定义随机变量的偏度系数为，大于零为正偏（右偏），小于零为负偏**

**定义随机变量的峰度系数为，大于零，并标准正态更尖峭，尾部更粗，小于零反之**

**次序统计量：将从小到大排序（对每个），那么有，其中为最小次序统计量**

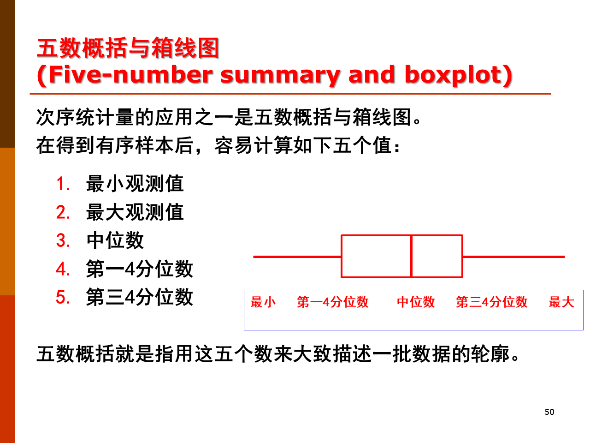
****

**样本极差**

**样本中程**

**定义样本中位数为：（n为奇数），（n为偶数）**

**同样定义p分位数为：（np不是整数），（np是整数）**

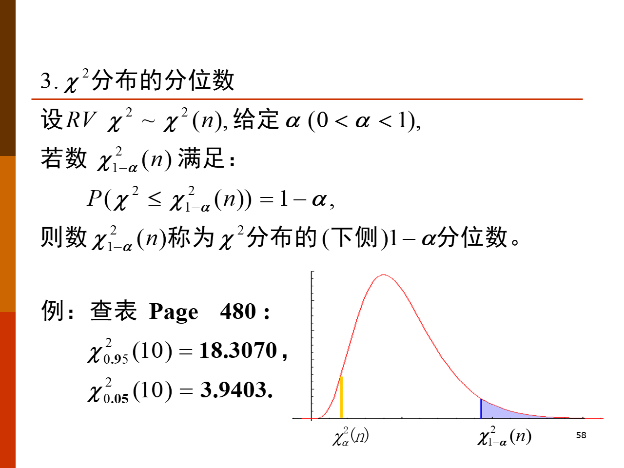
****

**三大抽样分布：分布，t分布，F分布**

**分布：设来自总体分布，称的分布为自由度为n的分布，记为**

**由于，那么由卷积公式，得到的密度为，其均值为，方差为**

**分位数：**

****

**t分布：现实问题中的方差常常未知，设，，二者独立，则称的分布为自由度为n的t分布，记为**

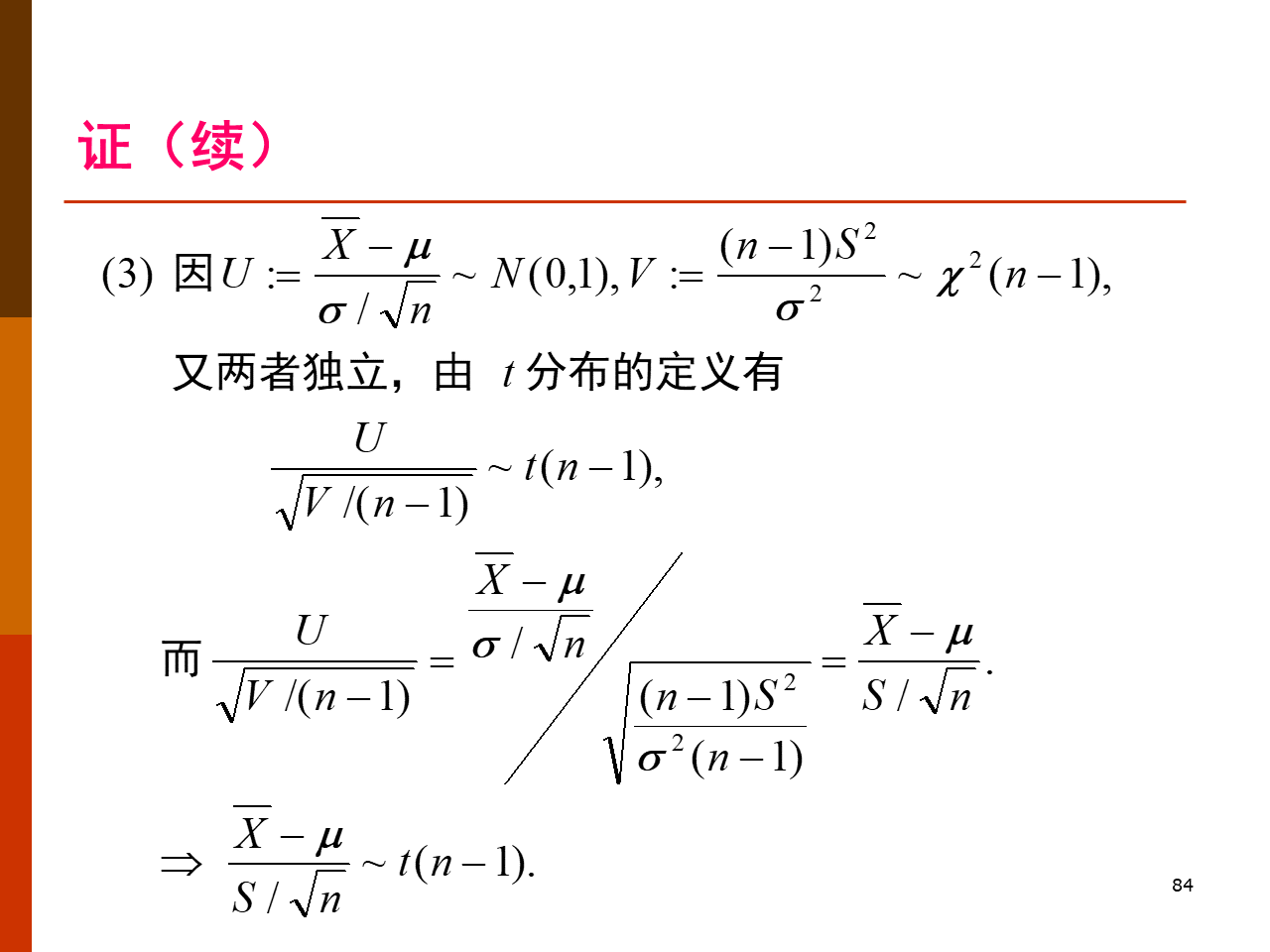
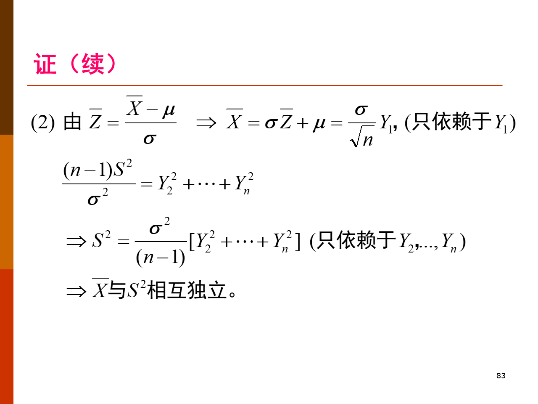
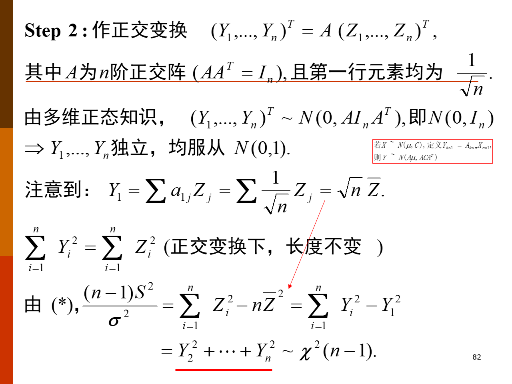
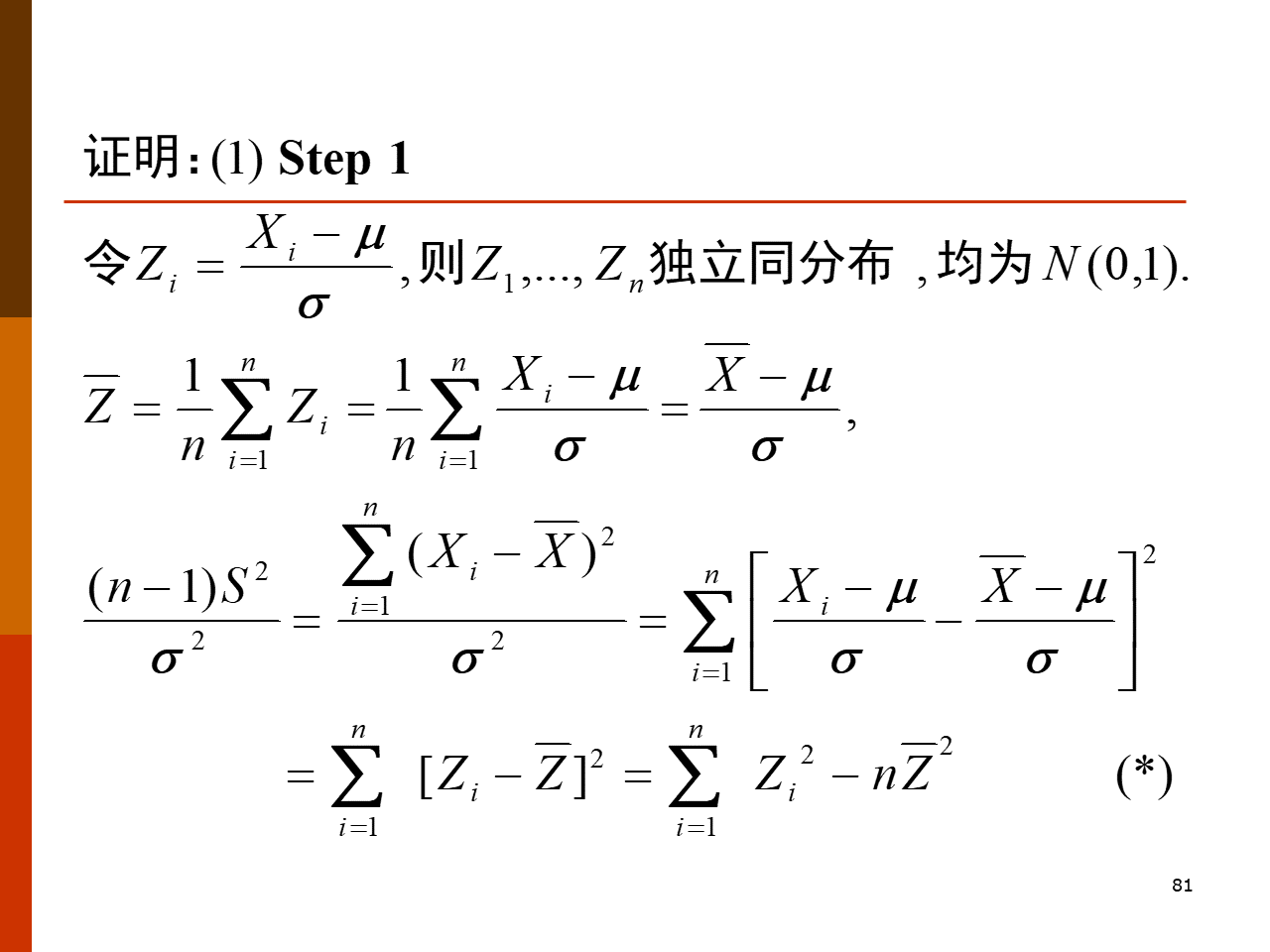
**t分布密度函数关于0对称，当n趋于无穷时，t分布的密度趋于的密度**

**自由度为1时，t分布为标准cauchy分布，期望不存在，当自由度大于1，期望存在且为零，n>2时，方差存在且为，n>=30时，t分布可用近似**

**F分布：设，，X与Y相互独立，则称为自由度为m和n的F分布，记为，m为第一自由度，n为第二自由度，F分布是取非负值的偏态分布**

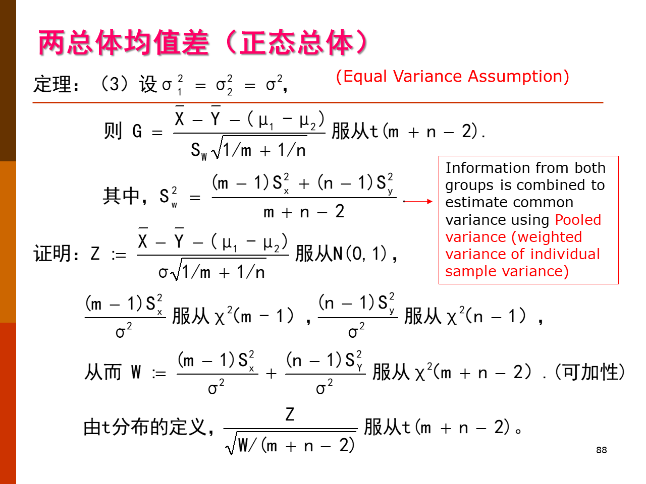
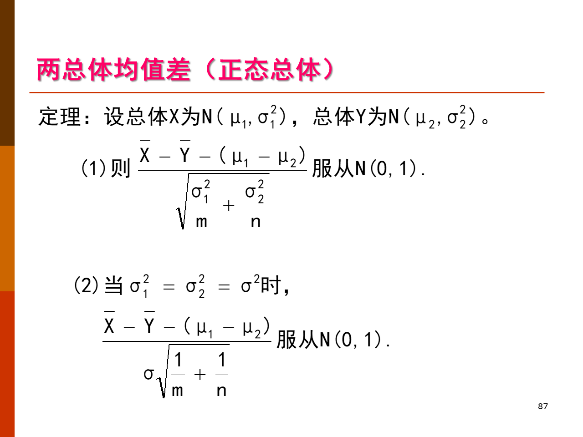
**F分布性质：①若，则②若，则**

**定理：设为来自正态总体的样本，那么其样本均值为，，那么有，与相互独立，**

****

**两个正态总体的均值差与方差比：**

**设在总体中选取m个样本，在总体中选取n个样本，那么其方差比有，另外有定理如下：**

****

**第六章 参数估计**

**参数：①分布中所含未知参数②未知参数的函数③分布的各种特征数**

**一般用表示参数，参数所有可能取值称为参数空间，用表示**

**设为总体的样本，利用样本给出参数的估计**

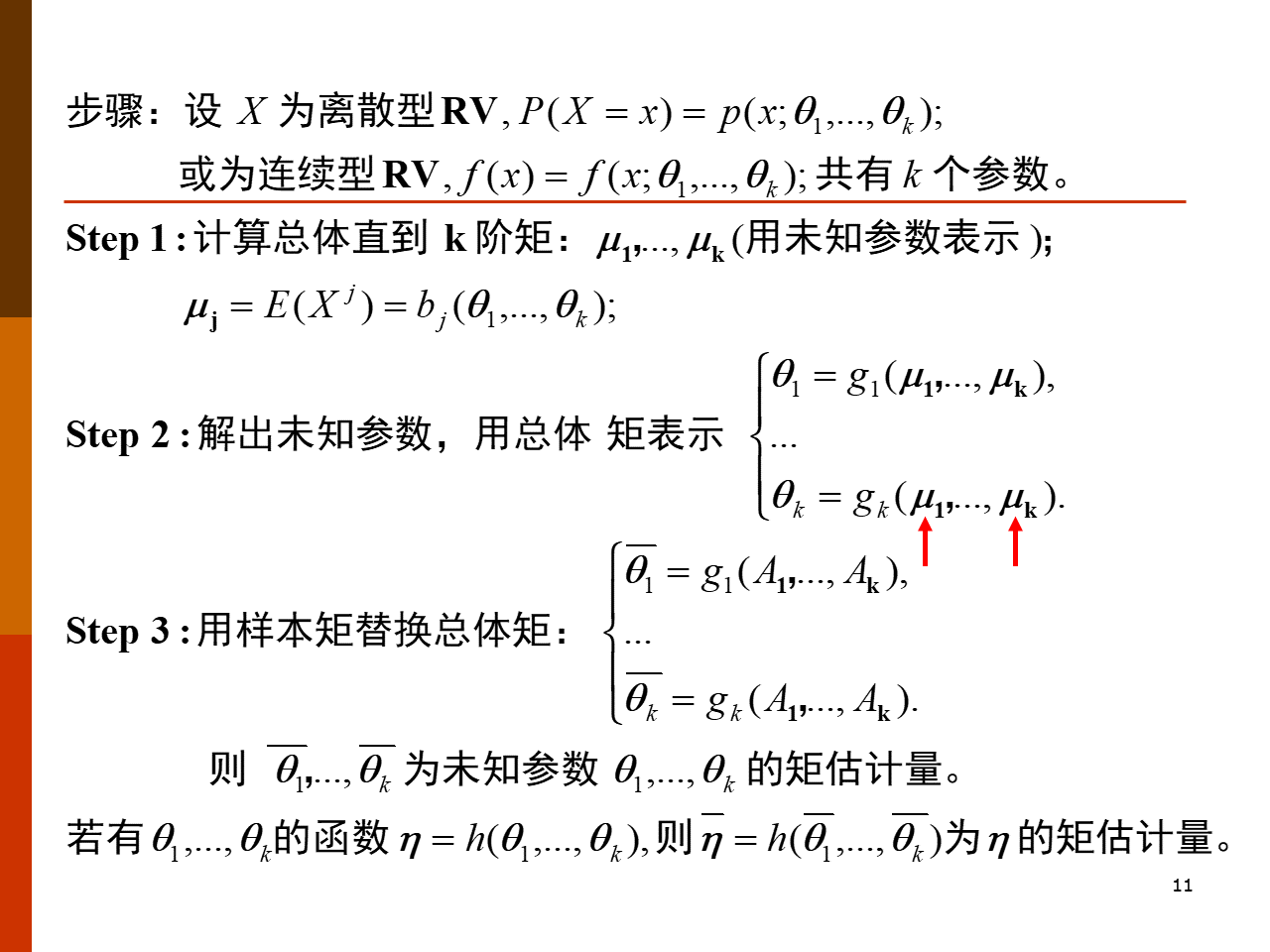
**参数估计形式：点估计（给出参数估计量）；区间估计（给出参数估计范围和参数可信度）**

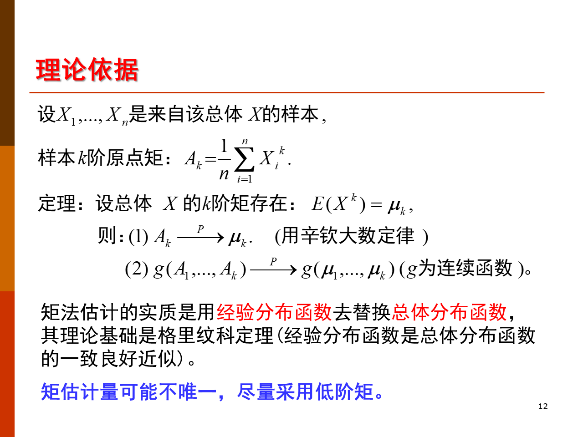
**需要构造估计量为估计量，使用观测值为参数估计值**

**点估计的两种方法：矩估计法；最大似然估计**

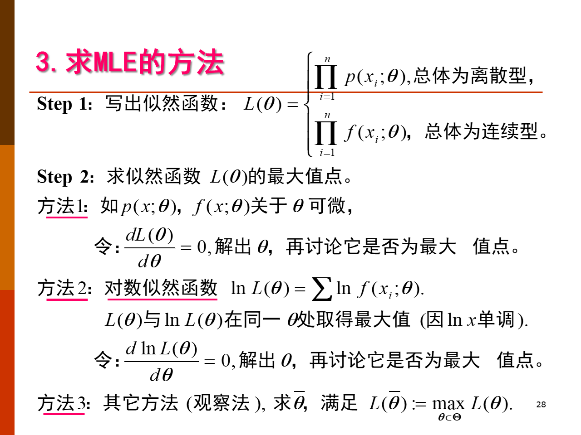
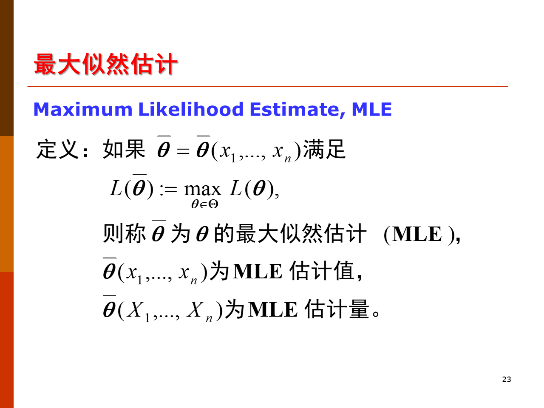
**矩估计法：**

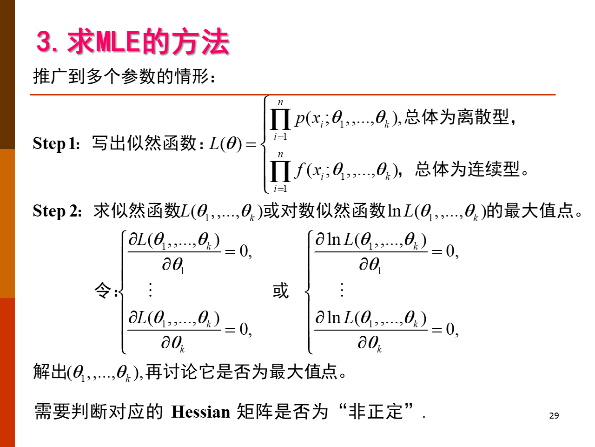
**替换原理：利用样本替换总体**

****

****

**最大似然估计：**

****

****

**点估计的评价标准：**

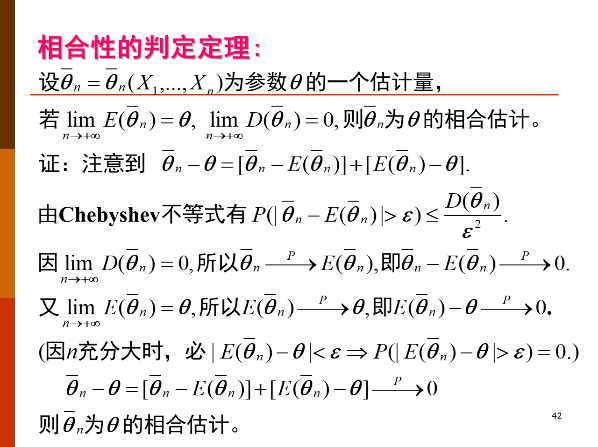
**估计量随着样品量不断增大而逼近参数真值，这就是相合性**

**相合估计：如果，其中，那么称为的相合估计**

**定理：设为相合估计，那么是的连续函数，则也是的相合估计**

**由此定理推断，矩估计一般具有相合性**

**相合性判定定理：若为参数的一个估计量，那么如果，同时，那么就是参数的相合估计**

****

**无偏性：若，则称其为无偏估计，否则为有偏估计**

**（称为系统误差，无偏性就是无系统误差）**

**无偏估计不具有不变性，也就是说，一般不是的无偏估计**

**有效性：若两个参数的无偏估计和，其中，那么称前者比后者更有效**

**均方误差：**

**P.59-P.64 未整理**

**区间估计：给出一个区间范围，希望它包含参数真值的可信度较高**

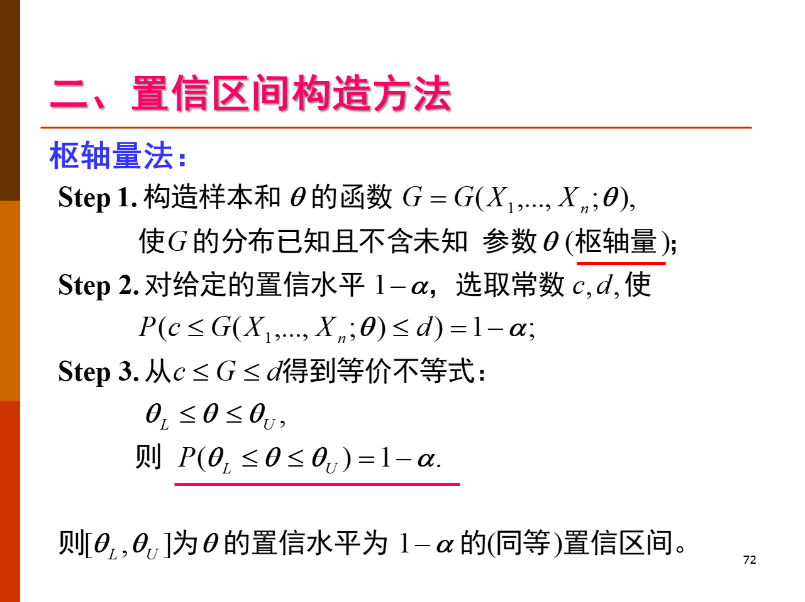
**设总体X服从正态分布，已知方差，未知均值，那么有均值的最大似然估计为MLE为，令，那么其落在某个区间的概率为，其中为可信度，解得可信度为的置信区间为**

**当方差未知，均值未知，令，那么均值的置信水平为的置信区间为**

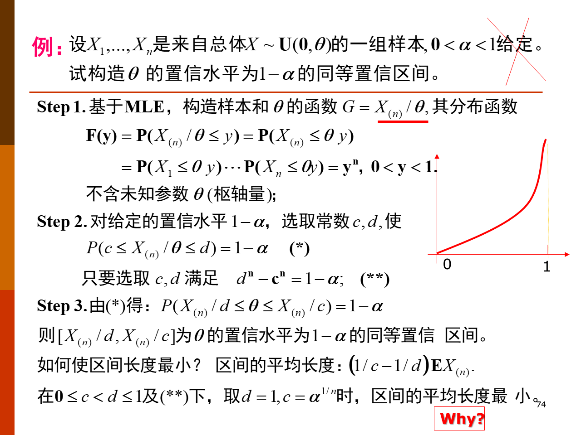
**区间估计的优良性准则：**

**可靠度，精度**

**置信区间构造方法：**

****

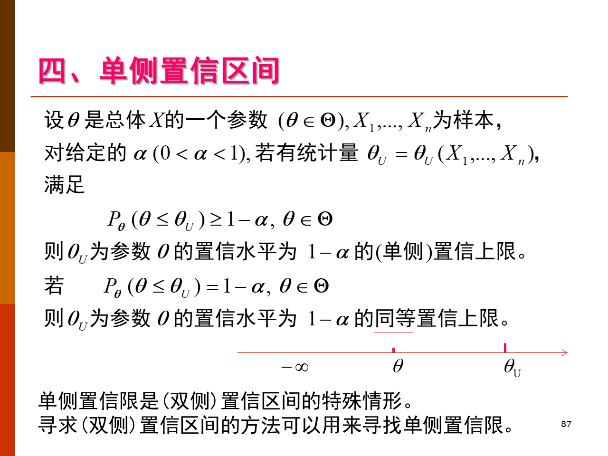
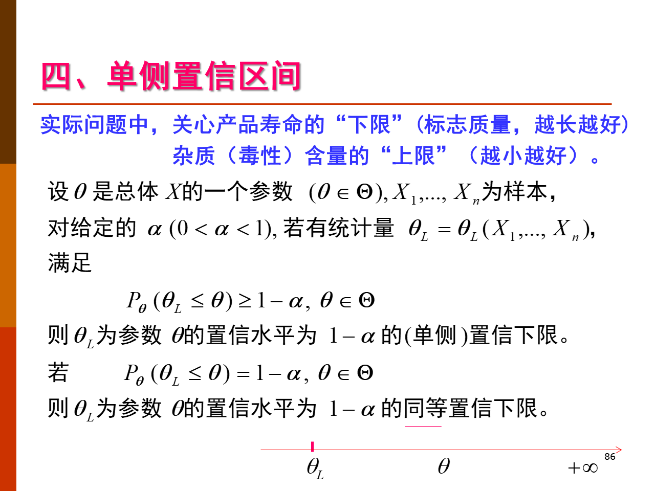
**c与d可能不唯一，尽可能选取使得置信区间平均长度最短者，常选择c与d使得两个尾部的概率各位，称为等尾置信区间（常用于偏态分布）**

****

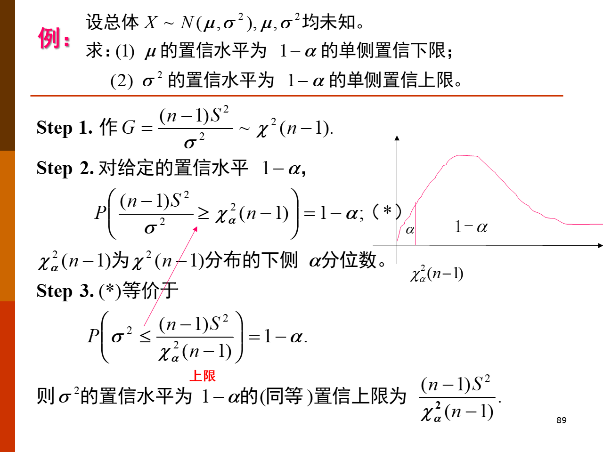
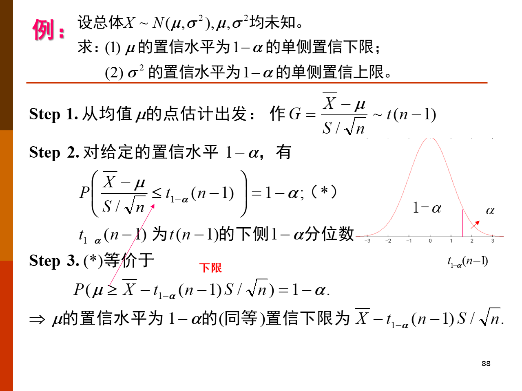
**方差的置信区间：均值未知时，取，那么其置信水平为的置信区间为（如果均值已知，那么取）**

**给定置信区间也可以确定置信水平**

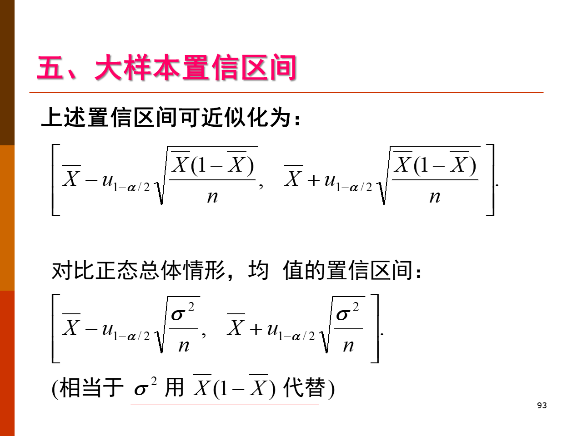
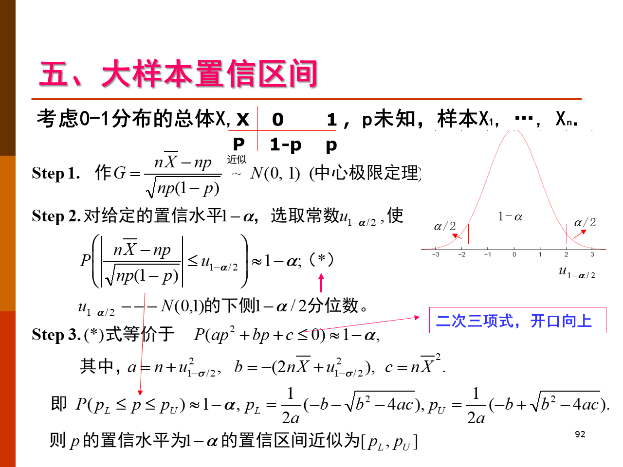
**单侧置信区间：实际问题中关心产品寿命的下限和杂质含量的上限**

****

**利用与置信区间相同的构造**

****

**大样本置信区间：**

****

**两个正态总体的置信区间：**

**P.97- 未整理**

**第七章 假设检验**

**若对参数未知，则采用参数估计方法**

**若对参数有怀疑，则用假设检验方法处理**

**假设检验的作用是用来推断样本与总体、样本与样本的差异是抽样误差还是本质误差**

**考试相关：**

**（前三个大数律会证明，辛钦大数律会用，两个中心极限定理）**

**不要求独立不同分布的中心极限定理**

**不要求特征函数**

**假设检验的分数不少**

**填空20个左右（一两道选择）**

**大题四道到五道（应该没有证明题）一道15到20分**