[随机变量及分布 5](#_Toc511065148)

[随机变量 10](#_Toc511065149)

[计算 14](#_Toc511065150)

[假设检验 16](#_Toc511065151)

## 事件与概率

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 事件  概率 | 随机试验E对应一个样本空间S()  样本空间：随机试验所有可能的结果组成的集合，样本空间的元素，称为样本点  随机事件：随机试验的某些可能的结果组成的集合，简称为事件，随机事件是样本空间的子集  概率：每一个随机事件对应一个实数P，称为该随机事件的概率，如随机事件A的概率P(A)，0≤P(A)≤1  概率为1的事件不一定是必然事件，P(A|B)=1⇏B⊂A | | | | | | | | | |
| 古典型概率 | | | | 几何形概率 | | | | 条件概率 | |
|  | | | |  | | | | 事件A发生的条件下，事件B发生的概率 | |
| 试验 | 独立重复试验 | | 某随机试验独立重复若干次，各次试验之间相互独立，同一事件在各个试验中出现的概率相同 | | | | | | | |
| n重伯努利试验 | | 独立重复试验，结果只有A和，独立重复n次  设P(A)=p，A发生k次的概率=(二项概率公式) | | | | | | | |
| 事件  关系 | 包含 | A发生时B必然发生  B⊂A | | | |  | |  | | |
| 相等 | B⊂A，A⊂B | | | |  | | P(A)=P(B) | | |
| 互斥  互不相容 | A和B不同时发生  A∩B=∅ | | | |  | | P(A∪B∪C)=P(A)+P(B)+P(C) | | |
| 相互独立 |  | | | |  | | P(AB)=P(A)P(B)  =  =  =  = | | |
| 互逆  对立 | A和B必有一个发生，且仅有一个发生  A∪B=Ω，A∩B=∅ | | | |  | | P(A)+P(B)=1 | | |
| 完备事件组 |  | | | |  | | 全概率：P()>0，则任意事件A： 贝叶斯：P()>0，P(A)>0，则任意事件A：P(|A)= | | |
| 事件  运算 | 积事件 | C=A∩B=AB  C={x│x∈A且x∈B}  =∅ A  AA=A | | | | C:\Users\ADMINI~1\AppData\Local\Temp\msohtmlclip1\02\clip_image008.png | | P(AB)=P(A)+P(B)-P(A+B)  P(AB)=P(A)-P(A-B)  P(AB)=P(A-)=P(B-) | | |
| 和事件 | C=A∪B=A+B  C={x│x∈A或x∈B}  A+A=A | | | | C:\Users\ADMINI~1\AppData\Local\Temp\msohtmlclip1\02\clip_image009.png | | P(A+B)=P(A-B)+P(B)  P(A+B)=P(A)+P(B)-P(AB)  P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC)  P(A+B)=1-P() | | |
| 差事件 | C=A-B=A  C={x|x∈A且x∉B} | | | | C:\Users\ADMINI~1\AppData\Local\Temp\msohtmlclip1\02\clip_image010.png | | P(A-B)=P(A+B)-P(B)  P(A-B)=P(A)-P(AB)  P(A-B)=P(A) | | |
| 逆事件 | C=  C={x│x∈S且x∉A} | | | | C:\Users\ADMINI~1\AppData\Local\Temp\msohtmlclip1\02\clip_image011.png | | P()=1-P(A) | | |
| 运算  律 | 交换律 | | | 结合律 | | | 分配率 | | | 德摩根律 |
| A+B=B+A  AB=BA | | | A+(B+C)=(A+B)+C  A(BC)=(AB)C | | | A+(BC)=(A+B)(B+C)  A(B+C)=(AB)+(BC) | | | = |

## 随机变量及分布

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 范围Ω | 性质 | 分布函数 | 数学期望 |
| 一维离散型 | x轴  点 | >0 =1 | F(x)=P{X≤x}=1-P{X>x} | = |
| 一维连续型 | x轴  线段 | =1  f(x)不一定连续  如果点x处连续，则(x)=f(x) | F(x)=P{X≤x}=1-P{X>x} 0≤F(x)≤1  F(-∞)=F(+∞)=1  F(x)必定右连续，=F() |  |
| 二维离散型 | 平面  点 | =1 | F(x,y)=P{X≤x,Y≤y}  0≤≤1 |  |
| 二维连续型 | 平面  区域 | f(x,y)≥0  = 1  f(x,y)=【X和Y相互独立】 | F(x,y)=P{X≤x,Y≤y}  0≤≤1  关于和关于均单调不减  关于和关于均右连续 |  |
| 二维  离散+连续  X离散  Y连续 | 平面  线段 | >0  ≥0  =1 | F(x,y)=P{X≤x,Y≤y}  (x|y)=P(X≤x,Y=y)  (y|x)=P(Y≤y,X=x) |  |
| 一维离散型函数  Y=g(X) | x轴  点 |  |  |  |
| 一维连续型函数  Y=g(X) |  |  |  |  |
| 二维离散型函数  Z=g(X,Y) |  |  |  |  |
| 二维连续型函数  Z=g(X,Y) |  |  |  |  |
| 二维离散+连续型函数  Z=g(X,Y) |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

对于任意给定的ε>0，P{y-ε<Y≤y+ε}>0，

则称此极限为在条件Y下X的条件分布，记作：或P，类似的可以定义

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 样本 |  |  |  |  |
| 样本均值 |  |  |  |  |
| 样本方差 |  | E()=D(X) |  |  |
| 样本标准差 |  |  |  |  |
| 样本k阶原点矩 |  |  |  |  |
| 样本k阶中心矩 |  |  |  |  |
| 统计量 |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | | |
| 一维离散型 | P{X=}=  X的每一个可能取值，对应一个概率  P=P{X=}+P{X=}+… | | |
| 一维连续型 | f(x) | X落在处，单位线段，概率 | |
| P{a≤X≤b} | P= | |
| P{a<X<b} | P=F()-F() | |
| F(x)在x=a连续时，P(x≤a)=P(x<a) | | |
| 二维离散型 | P{X=,Y= }= | | (X,Y)落在点(,)，的概率= |
| P{X= } | | 边缘~：P{X= }== |
| P{Y=} | | 边缘~：P{Y=}== |
| P{X=|Y=} | | 条件~：P{X=|Y=}= |
| P{Y=|X=} | | 条件~：P{Y=|X=}= |
|  | | P=P{X=a,Y=b }+P{X=c,Y=d }+… |
| 二维连续型 | f(x,y) | | (X,Y)落在处，单位面积，概率 |
| (x) | | 边缘~：(x)= ，计算时x看做常数 |
| (y) | | 边缘~：(y)= ，计算时y看做常数 |
| (x|y) | | 条件~：(x|y)= |
| (y|x) | | 条件~：(y|x)= |
| P{X∈D} | | 在平面中，找出D所包含的平面区域 P= |
| 二维离散+连续  X离散，Y连续 | P{X=}= | | X的每一个可能取值，对应一个概率，对应一条线段 |
| f(,y) | | (X,Y)落在(,y)处，单位线段，概率 |
| P{X∈D} | | 在平面中，找出D所包含的线段，如：  A：{X=，a1≤Y≤b1}  B：{X=，a2≤Y≤b2}  P=P{X=}·dy+P{X=}· |
| 一维离散型函数  Y=g(X) | 若，那么{Y=g}=+ | | |
| 一维连续型函数  Y=g(X) | 1. 确定x的范围：Ω，是一条线段  2. 令g(x)≤y，y看做常数，得到x的范围：D  3. y的不同取值范围，可能对应不同的D，则要分段讨论  4. F(y)=  f(y)=F’(y) | | |
| 二维离散型函数  Z=g(X,Y) | * X的分布律：P{X=}= * Y的分布律：{Y=}= * Z的分布律：P{Z=g(,)}= | | |
| 二维连续型函数  Z=g(X,Y) | 1. 确定(x,y)的范围：Ω，是一个平面区域 2. 令g(x,y)≤z，z看做常数，得到(x,y)的范围：D   F(z)=  f(z)= | | |
| 二维离散+连续型函数  Z=g(X,Y) | 1. 确定(x,y)的范围：Ω，是多条线段 2. 令g(x,y)≤z，z看做常数，得到(x,y)的范围：D   F(z)=  f(z)= | | |
| 样本 | 总体是一个随机变量X，作为某项数理统计研究  样本是一个随机变量序列：、…  任意一个样本，与总体X，同分布，同数字特征  **样本间相互独立** | | |
| 样本均值 | 与 ，相互独立  与 ，不互相独立 | | |
| 样本方差 |  | | |
| 样本标准差 |  | | |
| 样本k阶原点矩 |  | | |
| 样本k阶中心矩 |  | | |
| 统计量 | 统计量是一个随机变量  是关于一个样本的函数，T  如果是样本的一个样本值，  那么T()为统计量T的一个观测值  样本均值和样本方差都属于统计量 | | |
|  |  | | |

## 随机变量

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 二项分布  X~B(n,p)  0<p<1 | P(X=k)=  0≤k≤n，k∈N | E(X)=np  D(X)=np(1-p) | 试验成功率=p  独立重复试验n次  X：成功次数 | | |
| 0-1分布  X~B(1,p) | P(X=0)=p  P(X=1)=1-p | E(X)=p  D(X)=p(1-p) |  | | |
| 泊松分布  X~P(λ)  k≥0，λ>0 | P(X=k)=  P(X=0)= | E(X)=λ  D(X)=λ | |  |  | | --- | --- | | Z=  相互独立 | Z~P(n) | | | |
| 几何分布  X ~ GE(p)  0<p<1，k≥1 | P(X=k)=p | E(X)=  D(X)= | 试验成功率=p  独立地重复做该试验  X：第一次成功的次序 | | |
| 超几何分布  X~H(N,n,M) |  |  | N件产品中有M件次品  从中任取n件  X：n件产品中次品的 个数 | | |
| 均匀分布  X~U(a,b) | f(x)=，a<x<b  f(x)=0 ，其它 | E(X)=  D(X)= |  | | |
| 指数分布  X~E(λ)  λ>0 | f(x)=，x>0  f(x)=0 ，x≤0  C:\Users\ADMINI~1\AppData\Local\Temp\msohtmlclip1\02\clip_image002.jpg | E(X)=  D(X)= | P | | |
| 标准正态分布  Z~N(0,1) | C:\Users\ADMINI~1\AppData\Local\Temp\msohtmlclip1\02\clip_image002.gif | E(X)=0  D(X)=1 | aX+bY+c=·Z+c  X，Y相互独立，X，Y~N(0,1) | | |
|  | | |
| (n-1)~(n-1) | | |
| F(n-1,m-1) | | |
| 正态分布  X~N(μ,)，σ>0  X=σZ+μ，Z~N(0,1) |  | E(X)=μ  D(X)= | Z~N() | | |
|  | | |
| 二维均匀分布  (X,Y)~U(a,b) | A是平面有界区域G的面积 |  |  | | |
| χ2分布  X~  X=，~N(0,1) |  | E(X)=n  D(X)=2n |  | | |
|  | | |
| t分布  X~t(n)  X=，Z~N(0,1)，U~ |  |  | |  |  | | --- | --- | | n充分大时 | t(n)分布近似与N(0,1)分布 | | Z= | Z~F(1,n) | | | |
| F分布  X~F()  X=，U~，V~ |  |  |  |  |  |
|  |  |  |
|  |  |  |  | | |

|  |  |
| --- | --- |
|  | ：一个样本，来自总体  n：自由度 |
|  | 0<α<1 |

|  |  |
| --- | --- |
|  | X、Y相互独立  n：自由度 |
| 上分位点 |  |
| 上分位点 |  |

|  |  |
| --- | --- |
|  | (2)X、Y相互独立  (3)  (4)  (5)F服从自由度的F分布，分别称为第一自由度和第二自由度 |
| 上分位点 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 一个正态总体的抽样分布 | |  | | --- | | 设总体，、…是来自总体的样本，样本均值为，样本方差  那么 | |  | |  | |  | |  | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 两个正态总体的抽样分布 | |  | | --- | | 设总体，、…是来自总体X样本，样本均值为，样本方差  设总体，、…是来自总体Y样本，样本均值为，样本方差  那么 | |  | | 如果，则 | |  | |

|  |
| --- |
| t分布的概率密度f(x)是偶函数，即f(x)=f(-x) |
| P{T>c}=P{T<-c} |
| 由于t(n)分布的概率密度为偶函数，可知t分布的双侧a分为点，即  如图所示  显然 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 二维正态~ | (X,Y)~N(,,,,ρ)  X~N(,)、Y~N(,)  ρ是X和Y的相关系数 | X与Y相互独立的充分必要条件是ρ=0 |  |  |  | 服从二维正态，其中 |

## 计算

在计算一个随机变量的数字特征时，可以将它先转化为典型分布的随机变量

由于典型分布的数据特征是已知的，因此可以直接套用公式计算

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数学期望 | E(aX+bY+C)aE(X )+bE(Y )+C  E(XY)=E(X)·E(Y)+Cov(X,Y) | | | | | E(XY)=E(X)·E(Y) 【相互独立】  E()=D(X)+ | | | | E()= E( )= |
| 方差 | D(X)=E  D(X)=E()-  D(aX±bY+C)=D(X)+D(Y)±2abcov(X,Y) | | | | | D(aX±bY+C)=D(X)+D(Y) 【相互独立】 | | | |  |
| 协方差 | cov(X,Y)=cov(Y,X)=E(XY)-E(X)E(Y)    cov(X,Y)=· | | | |  | | | |  | |
| 相关系数 |  | | | | =0：X和Y不相关 | | | | X和Y相互独立  D(X)D(Y)=0  二维正态(X,Y)：X和Y相互独立 | |
| >0：正相关 | | | |  | |
| <0：负相关 | | | |  | |
| =±1 | | | | Y=aX+b a= ， a=E(Y)-E(X)a | |
| 矩 | k阶原点矩 | |  | | | | | | | |
| k阶中心矩 | |  | | | | | | | |
| m+n阶混合矩 | |  | | | | | | | |
| m+n阶混合中心矩 | |  | | | | | | | |
| 参数估计 | 估计量 | 样本：  统计量： 用来估计未知参数θ，称为估计量  估计量是随机变量  观测值称为估计值 | | | | | | | | |
| 无偏估计 | 是θ的无偏估计量：E()=θ  D()≤D()：称比更有效  依概率收敛于θ：称是θ的一致估计量 | | | | | | | | |
| 矩估计 | X的k阶原点矩 = ，  方程个数 = 未知参数个数  作为常数  解方程得未知参数的值：矩估计量  ………… | | | | | | | | |
| 最大似然估计 | 1. 构造似然函数：    * 离散型：    * 连续型： 2. L取最大值，对应的值：最大似然估计值    * 单调性相同    * 使用导数，求单调性或极点 | | | | | | | | |
| 区间估计 | 置信区间 | | 未知参数θ的，置信度为的，置信区间，()  是总体X的未知参数，是来自总体X的样本，对于给定的α  如果两个统计量满足P | | | | | | |
| 一个正态总体参数  区间估计 | | 总体X，是来自总体X的样本，是样本均值，是样本方差，  下表列出了μ和的1的置信区间 | | | | | | |
| 未知，已知 | | | |  | | |
| 未知，未知 | | | |  | | |
| 已知，未知 | | | |  | | |
| 两个正态总体参数  区间估计 | | 设总体，、…是来自总体X样本，样本均值为，样本方差  设总体，、…是来自总体Y样本，样本均值为，样本方差 | | | | | | |
| 未知  、已知 | | |  | | | |
| 未知  、未知 | | |  | | | |
|  | | |  | | | |

## 假设检验

|  |  |
| --- | --- |
| 假设检验问题 | 总体的分布完全未知或部分未知(知道分布函数的形式，但不知道其参数)  给出总体的一个或多个样本  提出关于所给样本的问题， |
| 提出假设 | 根据所给问题提出假设，即该问题所有可能的答案  如问题是判断题，那么答案只有两个（是或否），那么只需要提出两个假设  如问题是选择题，从n个选项中选择一个，那么答案就有n个，那么需要提出n个假设  在提出的假设中，(随意)选择一个，作为原假设(基本假设、零假设)，  那么其余的假设都称为备选假设(对立假设)，它们是原假设被拒绝后可供选择的假设  假设的性质：  参数假设：已知总体分布函数形式，对其中的未知参数，进行假设  非参数假设：直接对总体的分布函数进行假设  简单假设：假设提出后，总体分布确定  复合假设：假设提出后，总体分布不确定  双边假设：对未知参数进行范围假设，假设的范围是一个既有上界又有下界的范围  单边假设：对未知参数进行范围假设，假设的范围是一个只有上界的范围（右边假设）或只有下界的范围（左边假设） |
| 确定样本 | 假设一般都是针对样本的参数  因此要使用所求参数对应的样本 |
| 两类错误 | 决策过程就是选择某一个假设的过程，可能做出错误的决策  第一类错误：拒绝正确的假设，弃真  第二类错误：接受错误的假设，纳伪 |
| 显著性水平α | 在显著性检验中，只对第一类错误的概率加以控制，不考虑第二类错误  显著性水平α是允许犯第一类错误的概率，  对弃真的控制程度，一般α取：0.1、0.05、0.01、0.001 |
| 检验统计量 | 检验统计量是一个检验法则 |
| 拒绝域 | 当检验统计量取拒绝域中的值时，拒绝原假设  拒绝域的边界点称为临界点  按反第一类错误的概率等于 α，求出拒绝域W |
|  | 根据样本值计算检验统计量T的观测值t，当t∈W时，拒绝原假设，否则，接受原假设 |

正态总体参数的假设检验的

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 检验参数 | 情形 | 检验统计量，n是样本容量 |  | | |
| μ | 已知  Z检验法 |  | 原假设 | 备选假设 | 拒绝域 |
| μ= | μ≠ |  |
| μ≤ | μ＞ |  |
| μ≥ | μ＜ |  |
| μ | 未知  t检验法 |  | μ= | μ≠ |  |
| μ≤ | μ＞ |  |
| μ≥ | μ＜ |  |
|  | μ已知  检验法 |  |  |  |  |
|  |  |  |
|  |  |  |
|  | μ未知  检验法 |  |  |  |  |
|  |  |  |
|  |  |  |
|  | 已知  Z检验法 |  |  |  |  |
|  |  |  |
| ≥ | < |  |
|  | 未知  t检验法 |  |  |  |  |
|  |  |  |
| ≥ | < |  |
|  | 已知  F检验法 |  |  |  |  |
|  |  |  |
| ≥ | < |  |
|  | 未知  F检验法 |  |  |  |  |
|  |  |  |
| ≥ | < |  |
|  |  |  |  | | |