Salute to 张雪峰

目录

第一章 随机事件与概率

1.1 随机事件

1.1.1 样本空间

1. “观察”、“记录”一个个事件的过程，叫随机试验，简称试验，记作 E。

*试验的两个特征：1.试验可以在相同条件下重复地进行。2.试验可能出现的一切结果可事先预知，但不能事先确定每次试验会出现哪一个结果。

2.试验的每个可能结果称为样本点，记作ω；由试验的所有可能的结果所构成的集合，称为该试验的样本空间，记作Ω。

*即使是相同的随机试验，由于试验目的的不同，也可能对应不同的样本空间。

1.1.2 随机事件及其运算

在随机试验中，人们常常关心试验的某些结果是否出现，因此倾向于在被目的所标签了的总集合Ω的元素中，再以更细的分类标准来圈定划分新的小集合。

1.由样本点组成的集合称为随机事件，简称事件

仅由一个样本点组成的集合称为基本事件。

*相对于观察对象/试验目的来说，不能再分解的事件，即为基本事件，也就是样本点。【Atom 一词便是“a：不能”、“tom：切”，即“不能再切”的意思】

*一些基本事件构成复合事件，即事件。

*事件是样本空间的子集。

a.我们常常用大写字母 A、B、C 等标识事件。

b.若试验的结果(样本点)出现在事件 A 中，则称事件 A 发生。

c.由于每次试验结果必在Ω中，即事件Ω必然发生，则样本空间Ω又称为必然事件；不包含任何一个样本点的集合，即空集ϕ，在每次实验中必定没有样本点落在其中，即事件ϕ必不发生，则ϕ又称为不可能事件。

2.事件与事件间的关系与运算

由于事件可以用集合来表示，因此事件之间的关系和事件的运算，便可以按照集合论中，集合之间的关系和集合的运算来规定。下面给出这些关系在概率论中的提法，并根据 1.b. “事件发生”的含义，给出它们在概率论中的解释：

设 E 为一个随机试验，Ω为 E 的样本空间，A，B，A_i均为Ω的子集合，则：

a.关系:A⊂B 称为事件 B 包含事件 A，其概率含义为：A 发生必然导致 B 发生。

*若 A⊂B 且 A⊃B，则记为 A=B，称为事件 B 与事件 A 相等。

b.运算:A∪B={ω|ω∈A 或ω∈B}=A+B，称为事件 A 与事件 B 的和事件。其概率含义为：当且仅当 A、B 中至少一个发生时，事件 A∪B 才发生。

*无穷可列个事件和可表示为：$\bigcup_{i=1}^{n} A_i$。

c.运算:A∩B={ω|ω∈A 且ω∈B}=AB，称为事件 A 与事件 B 的积事件。其概率含义为：当且仅当 A、B 两者都发生时，事件 A∩B 才发生。

*可列无穷个事件的积事件可表示为：$\bigcap_{i=1}^{n} A_i$。

d.运算:$A\cap\bar{B}$={ω|ω∈A 且ω ∉B}=A−B，称为事件 A 与事件 B 的差事件。其概率含义为：当且仅当 A 发生且 B 不发生时，事件 $A\cap\bar{B}$才发生。

e.关系:若 A∩B=ϕ，则称事件 A 与事件 B 互不相容/互斥。其概率含义为：A 与 B 不能同时发生。

*若一组事件中任意两者间都互斥，则称该事件组为两两互斥事件组；基本事件是两两互斥事件组。

f.关系:若 A∩B=ϕ且 A∪B=Ω，则称事件 A 与事件 B 互为对立事件/互逆事件。其概率含义为：A、B 有且仅有一个发生。

*记 A 的对立事件为$\overline{A}$，则$\overline{A}$=B。

3.与集合论中的集合运算一样，事件之间的运算也满足

a.交换律：A∪B=B∪A、A∩B=B∩A。

b.结合律：A∪(B∪C)=(A∪B)∪C、A∩(B∩C)=(A∩B)∩C。(可推广)

c.分配律：A∪(B∩C)=(A∪B)∩(A∪C)、A∩(B∪C)=(A∩B)∪(A∩C)。(可推广)

d.德·摩根定律：$\overline{A\cup B}=\overline{A}\cap\overline{B}$、$\overline{A\cap B}=\overline{A}\cup\overline{B}$。(可推广)

*第四个中的第一个等式成立的原因：先证明其必要性(左推右)：若$\overline{A\cup B}$发生则$A\cup B$不发生，即 A、B 均不发生，所以$\overline{A}$、$\overline{B}$均发生，即$\overline{A}\cap\overline{B}$发生；反之，若$\overline{A}\cap\overline{B}$发生可同样推出$\overline{A\cup B}$发生。

****

4.有趣的课后题

①.将$\cup_{i=1}^{n}A_i$表示为 n 个两两互不相容事件的和。

思路：{A_1中的非$A_i(i\geq2)$部分}∪{A_2中的非$A_i(i\geq3)$部分}∪⋯∪{A_{n-1}中的非A_n部分}∪A_n。

解答：$\cup_{i=1}^{n}A_i=\cup_{j=1}^{n-1}[\cap_{i=j+1}^{n}\overline{A_i}\cap A_j]\cup A_n$。

链接：之后我们会用到 n=1 的情形：将 A 分为 AB(A 中的 B 部分)和 A-B(A 中的非 B 部分)这两个互斥(even 对立)的部分：A=(AB)+(A-B)=(A∩B)∪(A∩$\overline{B}$)=A∩(B∪$\overline{B}$)=A∩(Ω)=A。

②.把单位长度的一根细棒折成三段，三段细棒能构成一个三角形的概率。

思路：Ω={(x,y,z)|0<x,y,z<1,x+y+z=1}；A={(x,y,z)|x+y>z,y+z>x,x+z>y,x+y+z=1}。【注：A 本身应写作{(x,y,z)|x+y>z,y+z>x,x+z>y,0<x,y,z<1,x+y+z=1}】

解答：可见Ω，是个由坐标面的一卦限的放射状三棱锥空间将平面截下所得的，四面体的底面三角形，是个平面；而后者也是个放射状的三棱锥形空间截平面所得的四面体的底面的三个中点所连接而成的三角形，面积为其 1/4。【你可以通过

x+y+z= 1 将 x+y>z,y+z>x,x+z>y 转化为 x,y,z<0.5，再用这三个平面来截 x+y+z=1 这个平面】

1.2 随机事件的概率

1.2.1 频率与概率

a.在相同条件下，进行 n 次相同的试验中，事件 A 发生的次数n_A称为事件 A 发生的频数，比值$\frac{n_A}{n}$称为事件 A 发生的频率，记为$f_n(A)$。

易见频率$f_n(A)$具有如下基本性质：

(1).$0\le f_n(A)\le 1$

(2).$f_n(\Omega)=1$

(3).若A_1、$A_2 \cdots A_k$为两两互斥事件，则$f_n(A_1\cup A_2 \cdots\cup A_k)=f_n(A_1)+f_n(A_2)+\cdots+f_n(A_k)$。

b.设 E 为一个随机试验，Ω为 E 的样本空间，对 E 中的每个事件 A 赋予一个实数 P(A)，如果集合函数 P(·)满足：

(1).$P(A)\ge 0$ (非负性)

(2).$P(\Omega)=1$ (规范性)

(3).对于可列无穷个两两互斥事件A_1、$A_2 \cdots$，有 $P(A_1\cup A_2\cup\cdots)=P(A_1)+P(A_2)+\cdots$。

1.2.2 概率的运算性质

性质 1：$P(\phi)=0$

*证：令$A_i=\phi$，则$\cup_{i=1}^{\infty}A_i=\phi$，且$A_iA_j=\phi(i\ne j)$，即两两互斥，因此 $P(\phi)=P(\cup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)=\sum_{i=1}^{\infty}P(\phi)$，因此 $P(\phi)=0$。

性质 2：有限可加性

*设两两互斥事件A_1、$A_2\cdots A_n$，补充均为空集ϕ的$A_{n+1}\cdots$，那么A_1、$A_2\cdots A_{n+1}\cdots$也都两两互斥，且$\bigcup_{i=1}^{n}A_i=\bigcup_{i=1}^{\infty}A_i$，那么根据无穷可列事件的可列可加性 $P(\bigcup_{i=1}^{n}A_i)=P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)=\sum_{i=1}^{n}P(A_i)+\sum_{i=n+1}^{\infty}P(A_i)=\sum_{i=1}^{n}P(A_i)+\sum_{i=n+1}^{\infty}P(\phi)=\sum_{i=1}^{n}P(A_i)$。

性质 3：$P(\overline{A})=1-P(A)$

*由于 $A\cap\overline{A}=\phi$且 $A\cup\overline{A}=\Omega$，根据性质 2 和规范性，$1=P(\Omega)=P(A\cup\overline{A})=P(A)+P(\overline{A})$。

性质 4：$P(A\cup B)=P(A)+P(B)-P(AB)$ (加法公式)

*由于 $A\cap(B-A)=\phi$且 $A\cup(B-A)=A\cup B$，根据性质 2，$P(A\cup B)=P(A)+P(B-A)$。

又因 $BA\cap(B-A)=\phi$且 $BA\cup(B-A)=B$，则 $P(B)=P(BA)+P(B-A)$。两者联立即有之。

**$P(B)=P(BA)+P(B-A)$也可从另一个角度理解为 $P(B-A)=P(B)-P(BA)$；另外地，还将有 $P(BA)=P(B)-P(B-A)=P(A)-P(A-B)$。

***可用结合律和加法公式将加法公式推广，至三事件的情形：$P(A\cup B)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$；更多事件的运算也遵循：$+-+-\cdots$的规则。

性质 5：若 $A\subset B$，则 $P(A)\le P(B)$

*由于 $A\subset B$，则 $A\cap B=A$，则根据非负性，$P(B-A)=P(B)-P(BA)=P(B)-P(A)\ge 0$，因此有之。

性质 6：对任意事件 A，有 $P(A)\le 1$。

*根据性质 2，令 $B=\Omega$，再根据规范性，即有之。

****

1.一些或许会用到的公式及其推导

①.$P(A-B-C)=P(A)-P(AB)-P(AC)+P(ABC)$

思路：$P(A-B-C)=P(A\cap(\overline{B}\cap\overline{C}))=P(A\cap\overline{B\cup C})=P(A)-P(A-\overline{B\cup C})=P(A)-P(A\cap(B\cup C))=P(A)-P((A\cap B)\cup(A\cap C))=P(A)-[P(AB)+P(AC)-P(ABC)]=P(A)-P(AB)-P(AC)+P(ABC)$。

②.$P(A+B-C)=P(A)+P(B)-P(AB)-P(AC)-P(BC)+P(ABC)$

思路 1：$P(A+B-C)=P(A\cup B)-P(C\cap(A\cup B))=P(A)+P(B)-P(AB)-P((A\cap C)\cup(B\cap C))=P(A)+P(B)-P(AB)-P(AC)-P(BC)+P(ABC)$。

思路 2：$P(A+B-C)=P(A+B+C-C)=P(A\cup B\cup C)-P((A\cup B\cup C)\cap C)=P(A\cup B\cup C)-P(C)=P(A)+P(B)-P(AB)-P(AC)-P(BC)+P(ABC)$。

2.一道有意思的题

①.掷骰子 n 次，每次都不出现某个数字的概率。

思路：P(每次都不出现 1∪每次都不出现 2∪⋯)$=[P(\bar{1})+P(\bar{2})+\cdots+P(\bar{6})]-[P(\bar{1}\cap\bar{2})+P(\bar{1}\cap\bar{3})+\cdots]+[P(\bar{1}\cap\bar{2}\cap\bar{3})+\cdots]-\cdots=C_6^1\cdot(\frac{5}{6})^n-C_6^2\cdot(\frac{4}{6})^n+C_6^3\cdot(\frac{3}{6})^n-C_6^4\cdot(\frac{2}{6})^n+C_6^5\cdot(\frac{1}{6})^n-C_6^6\cdot(\frac{0}{6})^n$。

1.3 古典概型与几何概型

1.3.1 古典概型

a.如果一个随机试验 E 具有如下两个特点，则称试验 E 为古典概型：

(1).样本空间Ω包含有限个样本点，记$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$。

(2).每个样本点在一次试验中以相等的可能性出现，即 $P(\{\omega_1\})=P(\{\omega_2\})=\cdots=P(\{\omega_n\})=\frac{1}{n}$。

b.在古典概型中，如果事件 A 中含有n_A个样本点，则 $P(A)=\frac{n_A}{n}$称为古典概率。古典概率的计算涉及计数问题，而计数问题常用到一些排列组合公式，接下来给一些题目熟悉熟悉。

****

1.涉及排列组合的有趣题目

①.n 个球，每球均以 1/N 的概率放入 N(≥n)个盒子中(Maxwell-Boltzmann 统计)。

(1).某指定的 n 个盒子里各有一个球的概率：每个对应 n 的球均有 N 种 matching 盒子的方式，因此共 N^n 个样本点，相当于有放回地从 N 个球里取出 n 个(取 n 次)，即将球进盒子变为了盒子进球，两者均为将角标填空类型，而角标和空的角色互换了；分子应当为 n!。

(2).每个盒子至多有一个球的概率：只需将上一问的分子修改为A_N^n。

②.n 个人中没有两个人生日相同的概率：

(1).将 n 个球看作 n 个人，将 N 个盒子看作 365 天，结果即为①.(2).$\frac{A_{365}^n}{365^n}$。

(2).那么 n 个人中至少两个人生日相同的概率=$1-\frac{A_{365}^n}{365^n}$。

③.n 双相异的鞋分为 n 堆，恰好每堆成对的概率：$\frac{n!}{\left(\frac{(2n)!}{2^n}\right)}$

④.5 双中取 4 只，至少有 2 只配对的概率：$1-\frac{2^4\cdot C_5^4}{C_{10}^4}$

1.3.2 几何概型

a.如果一个随机试验 E 具有如下两个特点，则称试验 E 为几何概型：

(1).样本空间Ω是一维 or 二维 or 三维区域，并可以用长度、面积、体积来度量；且不管是全体样本点，还是我们所感兴趣的事件包含的样本点，都是无限的。

(2).样本点按照这种等可能性出现：设 A 为Ω的一个子集，P(A)与 A 的位置和形状无关，而只与 A 的度量(长度、面积、体积)成正比。【——这样其实可以推出每个点都是古典意义上的“等可能的”】

b.P(A)=$\frac{m(A)}{m(\Omega)}$称为几何概率。其中 m(·)表示以某种方式对·的度量，映射结果为·的长度、面积或体积。

****

1.几何概型中的经典题目

①.某码头只能容纳一只船，某日独立地来了两只船，在 T 内各时刻到来的可能性相等。它们分别需要停靠的时间为t_1、t_2，则有一只船需要等待进入码头的概率为：$1-\frac{\frac{(T-t_1)^2}{2}+\frac{(T-t_2)^2}{2}}{T^2}$。

设停t_1的船在 x 时刻到港，停t_2的船在 y 时刻到港。

法一：{y-x≤t_1}∩{y≥x}表示停t_2的船在 T 中需等待进入的到港时刻{y}们；而{x-y≤t_2}∩{x≥y}表示停t_1的船在 T 中需等待进入的到港时刻{x}们。取两者的并集即可。

法二：$\{y-x\geq t_1\}\cup\{y\leq x\}$表示停$t_2$的船在 T 中无需等待进入的到港时刻$\{y\}$们；而$\{x-y\geq t_2\}\cup\{x\leq y\}$表示停$t_1$的船在 T 中无需等待进入的到港时刻$\{x\}$们。取两者的交集后再用Ω来减之即可。

②.(Buffon 问题)平面上画有许多条平行线，相邻两条间距离为 a，向平面投掷一长度为 b(b<a)的针，则针与平行线相交的概率为：$\frac{2b}{\pi a}$。

设针的下端点必落于两个平行线之间，则针的下端点到它下面的那根线的距离为 y，设针尾指向针尖，相对于右指向的平行线，的逆时针偏转角为 x；可见 $y+b\sin x\geq a$ 是针与平行线相交的充分必要条件，其中 $0\leq x\leq\pi$(如果相交是由 $x>\pi$或 $x<0$ 导致的，那么请注意，这时的下端点已经变成了上端点了)、$0\leq y\leq a$。

那么$\Omega=\{(x,y)|0\leq x\leq\pi,\ 0\leq y\leq a\}$，而 $A=\{(x,y)|(x,y)\in\Omega,\ y\geq a-b\sin x\}$。则 $m(A)=\int_0^\pi dx\int_{a-b\sin x}^{a}dy=2b$，则 $P(A)=\frac{m(A)}{m(\Omega)}=\frac{2b}{\pi a}$。那么$\pi=\frac{2b}{a}\cdot\frac{1}{P(A)}\approx 2\cdot\frac{b}{a}\cdot\frac{n}{n_A}$，此种通过设计适当的随机试验而完成某种计算任务的途径，就是蒙特卡洛算法。

1.4 条件概率

a.定义当 P(A)>0 时(即 P(A)≠0 时)，$P(B|A)=\frac{P(BA)}{P(A)}$为在事件 A 发生的条件下，事件 B 发生的条件概率。

容易验证，条件概率 $P(\cdot|A)=\frac{P(\cdot A)}{P(A)}$符合概率定义中的三个条件，即：

(1).$P(\cdot|A)\geq 0$ (非负性)

(2).$P(\Omega|A)=1$ (规范性)

(3).对于可列无穷个两两互斥事件B_1、$B_2\cdots$，有 $P(B_1\cup B_2\cup\cdots|A)=P(B_1|A)+P(B_2|A)+\cdots$。

所以条件概率也是概率。它也具有概率的一些性质，如：$P(\overline{B}|A)=1-P(B|A)$。

*当 P(A),P(B)>0 时，有 $P(BA)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B)$，这叫做概率的乘法公式。它可推广到多个事件的情形：由于 $P(A)\geq P(BA)>0$，则 $P(CBA)=P(A)\cdot P(B|A)\cdot P(C|AB)$。对于 n 个事件，我们只需保证 $P(A_1A_2\cdots A_{n-1})>0$ 即可。

1.5 全概率公式与贝叶斯公式

a.设Ω为 E 的样本空间，B_1，B_2，…，B_n为一组事件，如果：

(1).$B_iB_j=\phi(i\neq j)$ (2).$B_1\cup B_2\cup\cdots\cup B_n=\Omega$，则称$B_1$，$B_2$，…，$B_n$为$\Omega$的一个划分，或称之为 E 的一个完备事件组。

b.设试验 E 的样本空间为Ω，B_1，B_2，… B_n为Ω的一个划分，且 $P(B_i)>0$，则：

全概率公式：$P(A)=P(B_1)\cdot P(A|B_1)+P(B_2)\cdot P(A|B_2)\cdots+P(B_n)\cdot P(A|B_n)$

*由于$B_iB_j=\phi$，因此$(AB_i)\cap(AB_j)=AB_iAB_j=AA\phi=\phi$，那么 $P(A)=P(A\Omega)=P(A\cap(B_1\cup B_2\cup\cdots\cup B_n))=P(AB_1\cup AB_2\cup\cdots\cup AB_n)=P(AB_1)+P(AB_2)+\cdots+P(AB_n)=P(B_1)\cdot P(A|B_1)+P(B_2)\cdot P(A|B_2)\cdots+P(B_n)\cdot P(A|B_n)$

c.设试验 E 的样本空间为Ω，B_1，B_2，… B_n为Ω的一个划分，且 $P(B_i)>0$，$P(A)>0$，则：

贝叶斯公式：$P(B_i|A)=\frac{P(B_i)\cdot P(A|B_i)}{\sum_{j=1}^{n}P(B_j)\cdot P(A|B_j)}$

*由条件概率中的乘法公式：$P(B_iA)=P(A)\cdot P(B_i|A)=P(B_i)\cdot P(A|B_i)$，则 $P(B_i|A)=\frac{P(B_i)\cdot P(A|B_i)}{P(A)}=\frac{P(B_i)\cdot P(A|B_i)}{\sum_{j=1}^{n}P(B_j)\cdot P(A|B_j)}$

*常用的两个形式：$P(\bar{B}|A)=\frac{P(\bar{B})\cdot P(A|\bar{B})}{P(\bar{B})\cdot P(A|\bar{B})+P(B)\cdot P(A|B)}$；$P(B|A)=\frac{P(B)\cdot P(A|B)}{P(B)\cdot P(A|B)+P(\bar{B})\cdot P(A|\bar{B})}$。

*在医疗诊断中，医生首先根据占人口百分比的资料，或者以往经验总结，来确定各个前提条件B_i所对应的先验概率 $P(B_i)$，再通过病理学或病历资料确定 $P(A|B_i)$，再通过贝叶斯公式得到各个病因B_i对病症 A 的出现起直接原因的可能性，即后验概率 $P(B_i|A)$，该概率越大，则患者表现为病症 A 的所患疾病越有可能是B_i。可以通过多个检查指标A_j，综合多组后验概率，看看每组A_j所对应的后验概率中，B_i对应的 $P(B_i|A_j)$是否都挺大，以此来诊断病情。

****

1.贝叶斯经典题目

①.某厂产品不合格率为 0.1%；一种仪器的误判率为 5%(即把合格判为不合格、把不合格判为合格的概率均为 5%)，厂方认为如果产品被判为不合格，但实际上产品也是不合格的概率低于 5%时，不能采用这种仪器，问此人发明的仪器能否被采用，若不能，得降低到多少。

我们设想维恩图的大圈被一根线分为真·合格与真·不合格，同样，维恩图的大圈也被一个圈分为内部和外部，内部为被判为不合格，外部为被判为合格；圈和线有两个交点，整个维恩图被分为 4 个空间(这里的圈也可以是与真·合格与真·不合格的分界线有一个交点的线，但这里为了显示地位上的不平等以示突出之用)。

于是真·合格为 P(B)=1-0.1%，真·不合格为 P($\overline{B}$)=0.1%；被判为不合格的圈内部，被线分为两部分，在真·合格中的部分，占真·合格的 P($\overline{A}$|B)=α%；在真·不合格中的部分，占真·不合格的 P($\overline{A}$|$\overline{B}$)=1-P(A|$\overline{B}$)=1-α%。这时

$$P(\overline{B}|\overline{A})=\frac{P(\overline{B})\cdot P\left(\overline{A}\middle|\overline{B}\right)}{P(\overline{B})\cdot P\left(\overline{A}\middle|\overline{B}\right)+P(B)\cdot P\left(\overline{A}\middle|B\right)}=\frac{0.1\%\cdot(1-\alpha\%)}{0.1\%\cdot(1-\alpha\%)+(1-0.1\%)\cdot\alpha\%}。$$

②.某种产品 50 件为一批，每批产品中没有次品的概率为 0.35，有 1,2,3,4 件次品的概率分别为 0.25、0.2、0.18、0.02.现从某批产品中随机抽取了 10 件，检查出一件次品，求该批产品中次品不超过 2 件的概率。

$$P(A)=\sum_{j=0}^{4}P(B_j)\cdot P\left(A\middle|B_j\right)=\sum_{j=0}^{4}P(B_j)\cdot\frac{C_{50-j}^{9}\cdot C_j^1}{C_{50}^{10}}；\ 而\ P(B_i|A)=\frac{\sum_{j=0}^{2}P(B_j)\cdot\frac{C_{50-j}^{9}\cdot C_j^1}{C_{50}^{10}}}{\sum_{j=0}^{4}P(B_j)\cdot\frac{C_{50-j}^{9}\cdot C_j^1}{C_{50}^{10}}}$$

1.6 事件的独立性

1.6.1 两个事件互相独立

a.如果 P(AB)=P(A)P(B)，则称 A，B 相互独立，简称 A，B 独立。

*判据之二：P(B|A)=P(B)或 P(A|B)=P(A)；判据之三：P(B|A)=P(B|$\overline{A}$)。

**若 P(A)、P(B)均>0(≠0)，则至少需要 A，B 有交集，A，B 才可能独立。——可见，非不可能事件的独立和互斥不会同时发生。

***若 A，B 独立，则 A 与$\overline{B}$、$\overline{A}$与 B、$\overline{A}$与$\overline{B}$也相互独立。证明：P(A$\overline{B}$)=P(A)-P(AB)=P(A)-P(A)P(B)=P(A)P($\overline{B}$)

****ϕ与任何事件都独立；ϕ与Ω既独立又互斥。

1.6.2 多个事件的相互独立

a.若 P(AB)=P(A)P(B)、P(AC)=P(A)P(C)、P(BC)=P(B)P(C)，则称 A，B，C 两两独立；若还有 P(ABC)=P(A)P(B)P(C)，则称 A，B，C 相互独立。

*两两独立不一定相互独立，相互独立一定两两独立；多个事件相互独立时，若将其中任意几个事件换成对立事件后，仍然相互独立。

*独立性在可靠性分析中有许多应用：对于一个元件，它能正常工作的概率称为元件的可靠性；由多个元件组成的系统，它能正常工作的概率称为系统的可靠性。若我们将“第 i 个元件能正常工作”这一事件用A_i表示，并记 P(A_i)=p_i，则：

(1).n 个元件串联的可靠性为：P($A_1A_2\cdots A_n$)=P(A_1)P(A_2)…P(A_i)

(2).n 个元件并联的可靠性为：P($A_1\cup A_2\cup\cdots\cup A_n$)=1-P($\overline{A_1\cup A_2\cup\cdots\cup A_n}$)=1-P($\overline{A_1}\,\overline{A_2}\cdots\overline{A_i}$)=1-P($\overline{A_1}$)P($\overline{A_2}$)…P($\overline{A_i}$)=1-$(1-p_1)(1-p_2)\cdots(1-p_i)$

对于复杂的系统，一般可分为若干串联或并联的子系统，进而求得整个系统的可靠性。

****

1.可靠性分析经典题目

①.2、3 号电阻串联后与 4 号电阻并联，之后整体再与 1 号电阻串联。

P=p[1-$(1-p^2)(1-p)$]=p[$p+p^2-p^3$]

②.1、2 号电阻串联，4、5 号电阻串联，两者并联；之后 3 号电阻左右两端分别接于 1、2 号电阻之间和 4、5 号电阻之间。

P=2·(1-p)·p[$p+p^2-p^3$]+p^2·[1-$(1-p)^2$]=$2p^5-5p^4+2p^3+2p^2$

其中，2·(1-p)·p 表示 2 断 5 顺的概率+5 断 2 顺的概率；[$p+p^2-p^3$]表示利用之前的结论：顺的那个电阻与类似于①中的 2、3、4 号电阻构成的系统串联；p^2表示 2 顺 5 顺的概率；[1-$(1-p)^2$]表示只需要 1、4 不同时断即可。

1.6.3 伯努利概型

只关心某个事件是否发生的试验称为伯努利试验。一个伯努利试验独立地做 n 次，n 个试验合称为 n 重伯努利试验。“一个伯努利试验独立地做 n 次”是指：P(A)=p 保持不变，各次试验的结果互相独立。有些试验的结果不只 2 个，若把结果中任何一部分当作关心的事件 A，其余部分为$\overline{A}$，则任何有 1 个以上结果的试验均可视为伯努利试验。在 n 重伯努利试验中，人们主要关心指定事件 A 发生指定次数的概率。

由于 n 个试验互相独立，并且指定的发生方式有C_n^k个且各个方式对应的事件之间均互斥，则$p_n(k)$=P(($AA\overline{A}\overline{A}\cdots$)∪($A\overline{A}A\overline{A}\cdots$)∪⋯)=$p^k\cdot(1-p)^{n-k}+p^k\cdot(1-p)^{n-k}+\cdots$=$C_n^k\cdot p^k\cdot(1-p)^{n-k}$；由于$\sum_{k=0}^{n}p_n(k)=(p+(1-p))^n$=1，因此这个概率称为二项概率。

****

1.章末习题

①.50 只铆钉随机取用于 10 个部件上，其中 3 个铆钉强度太弱，问它们均被装在同一个部件上的概率。

$$P=\frac{C_{10}^{1}\cdot C_{47}^{27}\cdot\frac{27!}{(3!)^9}}{C_{50}^{30}\cdot\frac{30!}{(3!)^{10}}}$$

②.每条蚕产 n 个卵的概率为$p_n=\frac{\lambda^n}{n!}\cdot e^{-\lambda}$；每个卵变为成虫的概率相互独立，为 p。

(1).每条蚕养出 k 条成虫的概率：P=$\sum_{n=k}^{\infty}C_n^k\cdot p^k\cdot(1-p)^{n-k}\cdot p_n=e^{-\lambda}\cdot\frac{p^k}{(1-p)^k}\cdot\sum_{n=k}^{\infty}\frac{C_n^k}{n!}(1-p)^n\cdot\lambda^n=e^{-\lambda}\cdot\frac{p^k}{(1-p)^k\cdot k!}\cdot\sum_{n=k}^{\infty}\frac{(1-p)^n\cdot\lambda^n}{(n-k)!}=e^{-\lambda}\cdot\frac{p^k}{(1-p)^k\cdot k!}\cdot\sum_{N=0}^{\infty}\frac{(1-p)^{N+k}\cdot\lambda^{N+k}}{N!}=e^{-\lambda}\cdot\frac{(\lambda p)^k}{k!}\cdot\sum_{N=0}^{\infty}\frac{(1-p)^N\cdot\lambda^N}{N!}=e^{-\lambda}\cdot\frac{(\lambda p)^k}{k!}\cdot e^{\lambda(1-p)}=\frac{(\lambda p)^k}{k!}\cdot e^{-\lambda p}$，这个结果相当于将$p_n$中的 $n\to k$、$\lambda\to\lambda p$。

(2).每条蚕养了 k 条成虫，其产了 n 个卵的概率：$\frac{C_n^k\cdot p^k\cdot(1-p)^{n-k}\cdot p_n}{\frac{(\lambda p)^k}{k!}\cdot e^{-\lambda p}}=\frac{C_n^k\cdot k!}{n!}\cdot\frac{\lambda^n}{\lambda^k}\cdot(1-p)^{n-k}\cdot e^{-\lambda+\lambda p}=\frac{[\lambda(1-p)]^{n-k}}{(n-k)!}\cdot e^{-\lambda(1-p)}$

③.A、B 两人轮流射击，每次每人一枪，次序为 ABAB⋯，直到击中 2 枪停止，两人击中概率均为 p；求击中的两枪是由同一个人射中的概率 P。【下面有三个有趣式子】

p_1=P(两枪就能击中的概率)=$\frac{\sum_{i=0}^{\infty}C_{i+2}^{2}\cdot p^2(1-p)^i}{\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}C_{i+j}^{j}\cdot p^j(1-p)^i}$，而分子可以这样算：$\sum_{i=0}^{\infty}C_{i+2}^{2}\cdot p^2(1-p)^i=p^2\cdot\sum_{i=0}^{\infty}(C_{i+1}^{2}+C_{i+1}^{1})\cdot(1-p)^i=p^2\cdot\sum_{i=0}^{\infty}(C_2^2+\sum_{j=2}^{i+1}C_j^1)\cdot$

$(1-p)^i=p^2\cdot\sum_{i=0}^{\infty}\sum_{j=1}^{i+1}C_j^1(1-p)^i=p^2\cdot\sum_{j=1}^{\infty}\sum_{i=j-1}^{\infty}C_j^1(1-p)^i=p^2\cdot\sum_{j=1}^{\infty}C_j^1\sum_{i=j-1}^{\infty}(1-p)^i===p^2\cdot\sum_{j=1}^{\infty}C_j^1\frac{(1-p)^{j-1}}{1-(1-p)}=p^2\cdot\sum_{j=1}^{\infty}\sum_{i=1}^{j}\frac{(1-p)^{j-1}}{p}=p^2\cdot\sum_{i=1}^{\infty}\sum_{j=i}^{\infty}\frac{(1-p)^{j-1}}{p}=p^2\cdot\sum_{i=1}^{\infty}\frac{(1-p)^{i-1}}{p^2}=\sum_{i=1}^{\infty}(1-p)^{i-1}=\frac{1}{p}>1$

而分母$\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}C_{i+j}^j\cdot p^j(1-p)^i=\sum_{n=0}^{\infty}\sum_{i+j=n}C_{i+j}^j\cdot p^j(1-p)^i=\sum_{n=0}^{\infty}\sum_{j=0}^{n}C_n^j\cdot p^j\cdot(1-p)^{n-j}=\sum_{n=0}^{\infty}1=\infty$；分母若硬算的话不好算，我们可以试一下$\sum_{i=0}^{\infty}C_{i+j}^j\cdot p^j(1-p)^i$中 j=3 的情景：$\sum_{i=0}^{\infty}C_{i+3}^3\cdot p^3(1-p)^i=\sum_{i=0}^{\infty}(C_3^3+\sum_{j=3}^{i+2}C_j^2)\cdot p^3(1-p)^i=\sum_{i=0}^{\infty}\sum_{j=2}^{i+2}C_j^2\cdot p^3(1-p)^i=\sum_{j=2}^{\infty}\sum_{i=j-2}^{\infty}C_j^2\cdot p^3(1-p)^i=\sum_{j=2}^{\infty}C_j^2\cdot p^3\frac{(1-p)^{j-2}}{1-(1-p)}=\sum_{j=2}^{\infty}\sum_{i=1}^{j-1}p^2(1-p)^{j-2}\cdot i=\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}p^2(1-p)^{j-2}\cdot i===\sum_{i=1}^{\infty}p^2\frac{(1-p)^{i-1}}{p}\cdot i=\sum_{i=1}^{\infty}\sum_{j=1}^{i}p(1-p)^{i-1}=\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}p(1-p)^{i-1}=\sum_{j=1}^{\infty}(1-p)^{j-1}=\frac{1}{p}$，其实从红等号开始，就已经重蹈覆辙了。

j=3 的另一种快速算法，利用 j=2：$\sum_{i=0}^{\infty}C_{i+3}^3\cdot p^3(1-p)^i=\sum_{i=0}^{\infty}(C_{i+2}^3+C_{i+2}^2)\cdot p^3(1-p)^i=p\sum_{i=0}^{\infty}C_{i+2}^2\cdot p^2(1-p)^i+\sum_{i=1}^{\infty}C_{i+2}^3\cdot p^3(1-p)^i=p\cdot\frac{1}{p}+\sum_{i=0}^{\infty}C_{i+3}^3\cdot p^3(1-p)^{i+1}=1+(1-p)\sum_{i=0}^{\infty}C_{i+3}^3\cdot p^3(1-p)^i$，即有：P=1+$(1-p)$P，于是 P=$\frac{1}{p}$；可见之前(j=1)和之后每一 j 级均为$\frac{1}{p}$。

p_2=P(击中两枪且最后一枪为击中)=$\sum_{n=2}^{\infty}(n-1)\cdot p^2(1-p)^{n-2}=\sum_{i=2}^{\infty}\sum_{j=1}^{i-1}p^2(1-p)^{i-2}=\sum_{j=1}^{\infty}\sum_{i=j+1}^{\infty}p^2(1-p)^{i-2}=\sum_{j=1}^{\infty}p(1-p)^{j-1}=1$。

p_3=P(击中两枪的是同一人且最后一枪为击中)=$\sum_{n=1}^{\infty}n\cdot p^2(1-p)^{2n-1}+\sum_{n=1}^{\infty}n\cdot p^2(1-p)^{2n}=\sum_{n=1}^{\infty}n\cdot p^2((1-p)^{2n-1}+(1-p)^{2n})=\sum_{i=1}^{\infty}\sum_{j=1}^{i}p^2((1-p)^{2i-1}+(1-p)^{2i})=\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}p^2((1-p)^{2i-1}+(1-p)^{2i})=\sum_{j=1}^{\infty}\sum_{i=2j-1}^{\infty}p^2(1-p)^i=\sum_{j=1}^{\infty}p(1-p)^{2j-1}=p\frac{(1-p)}{1-(1-p)^2}=\frac{1-p}{2-p}$

该题答案即为p_3/p_2。

第二章 随机变量 r.v.(random.variety.)及其分布

2.1 事件的独立性

a.每个试验结果(样本点)ω对应着一个实数 X(ω)，相当于在样本空间Ω中定义了一个实值函数。反过来，每一个实数 X 都是对“具有属性 X”的一类随机事件的标识。随机事件发生的概率即为标识随机事件的实数 X 发生的概率，因而 X 也具有随机性。这种取值随机的变量就称为随机变量。

*随机变量通常用大写字母或者希腊字母表示，而它们的取值通常用英文小写字母表示。相当于随机变量名字为x_0，其取值是 x；变量 x，其取值是 a；变量 a，其取值是 3。

a.a.当试验结果取值为可列个时，对应的随机变量称为离散型随机变量；否则称为非离散型随机变量，其中重要而常见的一种是连续型随机变量。

b.能够描述随机变量的全部可能取值、它取这些值所对应的概率，这两者的对应关系的函数、表、图，称为概率分布律(这些概率关系分别对应地称为概率函数、概率分布表、概率分布图)，简称概率分布。

2.2 离散型随机变量及其分布

2.2.1 离散型随机变量的分布律

a.离散型 r.v.的概率函数的一般形式为：$f(x_i)=P\{X=x_i\}=p_i$，其中“$\{X=x_i\}$”表示“取值为x_i的随机变量 X 所标识(对应)的事件”

离散型 r.v.的概率分布律具有如下性质：$p_i \geq 0$、$\sum_i p_i = 1$。

2.2.2 常见的离散型随机变量及其分布

1.0-1 变量及其分布

只有两个可能的取值的随机变量叫两点变量，它所服从的分布为两点分布。两点变量中最常见的是取值为 0 和 1 的 0-1 变量，其分布律为：P{X=k}=$p^k(1-p)^{1-k}$，k=0,1。——这称为参数为 p 的 0-1 分布。

*0-1 变量常用来描述伯努利试验的结果：{X=1}=事件 A，{X=0}=事件$\overline{A}$，故也称 0-1 分布为伯努利分布。

2.二项变量及其分布

之前 n 重伯努利试验中，事件 A 发生 k 次的概率$p_n(k)=C_n^k \cdot p^k \cdot (1-p)^{n-k}$，由于它恰好是二项式$(p+(1-p))^n$展开式中的一般项，因此表示发生次数 k 的随机变量 X 叫二项变量。其分布律为：P{X=k}=$C_n^k \cdot p^k \cdot (1-p)^{n-k}$，k=0,1,2,⋯,n。——这称为参数为 n,p 的二项分布，记为 X~B(n,p)。

*当 n=1 时，二项变量即为 0-1 变量。

*P{X=x}在 x=[(n+1)p]处达到最大值：$\frac{P\{X=x+1\}}{P\{X=x\}}=\frac{n-x}{x+1}\cdot\frac{p}{1-p}$，若 x 从小取到大，过程中此值刚刚≤1 时，此时的 x 值对应的函数值最大；若此时比值恰=1，则 x 与 x+1 对应的函数值相等；则我们要求满足$\frac{n-x}{x+1}\cdot\frac{p}{1-p}\leq 1$的最小 x 值(或者说$\frac{n-x}{x+1}\cdot\frac{p}{1-p}>1$的最大 x 值+1)，即有$(n-x)p\leq(x+1)(1-p)$，$np\leq x-p+1$，得到$x\geq(n+1)p-1$，由于 x 为大于等于它的最小整数，也就是说我们要向右取整；然而数学中，取整符号是向左取整的——一个办法是，将$(n+1)p-1$加 1 后再向左取整，即有$[(n+1)p]$——但是这种方法只有当$(n+1)p-1$不是整数时，才起向右取整的功效；若$(n+1)p-1$是整数，则$[(n+1)p]$得到的是$(n+1)p$，而不是所期望的$(n+1)p-1$，其对应的函数值为我们想要的值的下一个值。——不过此时幸好$(n+1)p$所对应的函数值与$(n+1)p-1$这个最大值处所在的函数值是一样的(因为此时整数 x=整数$(n+1)p-1$，$\frac{P\{X=x+1\}}{P\{X=x\}}=\frac{n-x}{x+1}\cdot\frac{p}{1-p}=1$)，因此我们可以直接选用$[(n+1)p]$来求取最大值；另一种方法是，$-[-((n+1)p-1)]$，即将$(n+1)p-1$关于原点对称后，向左取整，再关于原点对称变换一次。这样得到的便是准确的值了，不过它看上去不好看而没被采用。

*使得 P{X=x}达到最大值的二项变量 X 的取值称为最可能数。

3.泊松变量及其分布

当 n 很大时，很难给出二项变量分布律的具体结果，而实际中又经常碰到 n 很大的情形，因此我们开发了二项分布的泊松逼近方法：记$\lambda_n=n\cdot p_n$，则 $\lim\limits_{n\to\infty} C_n^k\cdot p^k\cdot(1-p)^{n-k}=\lim\limits_{n\to\infty}\frac{n(n-1)\cdots(n-k+1)}{k!}\cdot(\frac{\lambda_n}{n})^k\cdot(1-\frac{\lambda_n}{n})^{n-k}=\lim\limits_{n\to\infty}\frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{\lambda_n^k}{k!}\cdot(1-\frac{\lambda_n}{n})^{-k}\cdot(1-\frac{\lambda_n}{n})^{n}=\frac{\lambda_n^k}{k!}\cdot\lim\limits_{n\to\infty}1\cdot(1-\frac{1}{n})\cdots\left(1-\frac{k-1}{n}\right)\cdot(1-\frac{\lambda_n}{n})^{-k}\cdot[(1-\frac{\lambda_n}{n})^{\frac{n}{\lambda_n}}]^{\lambda_n}$，由于即使 n 很大时，$p_n$也很小，保证$\lambda_n$近似是个常数，此时又因 k 为常数，则$1\cdot(1-\frac{1}{n})\cdots\left(1-\frac{k-1}{n}\right)\cdot(1-\frac{\lambda_n}{n})^{-k}\approx1$，于是原式变为：$\frac{\lambda_n^k}{k!}\cdot\lim\limits_{n\to\infty}[(1-\frac{\lambda_n}{n})^{\frac{n}{\lambda_n}}]^{\lambda_n}=\frac{\lambda_n^k}{k!}\cdot e^{-\lambda_n}$。

于是二项分布的极限情形 $P\{X=k\}=\frac{\lambda^k}{k!}\cdot e^{-\lambda}$，k=0,1,2,⋯，$\lambda>0$，称为参数为$\lambda$的泊松分布，并记 X~P(λ)；其中 X 称为泊松变量。

*实际应用中，n≥100，λ=np(或者 n(1-p))<10 时，就可以用泊松分布近似计算二项分布。【n 很大，p 很大或很小均可；若 n 很大，p 不大也不小，则此时二项分布近似于正态分布而不是泊松分布】

*事件流服从泊松分布的条件：(1).平稳性：任意时间区间内，事件发生 k 次的概率只与区间长度有关，与区间端点无关。(2).无后效性：在不相重叠的时间段内，事件的发生相互独立。(3).普通性：若时间区间充分小，事件发生 2 次的概率忽略不计。

4.超几何变量及其分布

有 N 件产品，其中 M 件为次品，从中抽出 n 件，则抽出的次品数 X 的分布律为：$P\{X=k\}=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}$，k=max{0,n-(N-M)},前者+1,⋯,min{n,M}。称为参数为 N,M,n 的超几何分布，并记为：X~H(N,M,n)；X 称为超几何变量。

*当总量 N 较大，且抽取量 n 比较小时，不放回抽取可当作有放回抽取处理，这是因为：

$$\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}=\frac{M!}{k!(M-k)!}\cdot\frac{(N-M)!}{(n-k)!(N-M-n+k)!}\cdot\frac{n!(N-n)!}{N!}=\frac{n!}{k!(n-k)!}\cdot\frac{M(M-1)\cdots(M-k+1)}{1}\cdot\frac{(N-M)(N-M-1)\cdots(N-M-n+k+1)}{1}\cdot\frac{1}{N(N-1)\cdots(N-n+1)}=C_n^k\cdot\frac{M(M-1)\cdots(M-k+1)}{N^k}\cdot\frac{(N-M)(N-M-1)\cdots(N-M-n+k+1)}{N^{n-k}}\cdot\frac{N^n}{N(N-1)\cdots(N-n+1)}=C_n^k\cdot\frac{M}{N}\left(\frac{M}{N}-\frac{1}{N}\right)\cdots\left(\frac{M}{N}-\frac{k-1}{N}\right)\cdot\left(1-\frac{M}{N}\right)\left(1-\frac{M}{N}-\frac{1}{N}\right)\cdots\left(1-\frac{M}{N}-\frac{n-k-1}{N}\right)\cdot\frac{1}{1\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}\approx C_n^k\cdot\left(\frac{M}{N}\right)^k\cdot\left(1-\frac{M}{N}\right)^{n-k}$$

5.几何变量及其分布

重复独立地做伯努利试验，直到事件 A 首次发生为止，已进行的试验次数 X 的分布律：P{X=k}=$(1-p)^{k-1}p$，k=1,2,⋯。称为参数为 p 的几何分布，并记 X~G(p)。X 称为几何变量。

*几何分布具有无记忆性：P{X=k}=$\frac{P\{X=m+k\}}{P\{X=m\}\cdot\frac{1-p}{p}}$=$(1-p)^{k-1}p$，k=1,2,⋯。即如果在前 m 次试验中事件 A 未发生，那么直到事件 A 首次发生为止，再进行的试验数 X 的分布仍然服从几何分布，与前面的试验数 m 无关。

****

6.有意思的两道习题

①.某实验室有 10 台相互独立运行的机器，发生故障的概率均为 0.03。一个机器需要一个技师处理，问配备多少技师可以保证设备发生故障时不能及时维修的概率<0.05。

P(设备发生故障不能及时维修，技师数为 k-1)=P(k 台及以上机器同时故障)=$C_{10}^{10}\cdot 0.03^{10}\cdot 0.97^{0}$+⋯+$C_{10}^{k}\cdot 0.03^{k}\cdot 0.97^{10-k}$=1-P(k-1 台及以下机器同时故障)=1 − ($C_{10}^{0}\cdot 0.03^{0}\cdot 0.97^{10}$+⋯+$C_{10}^{k-1}\cdot 0.03^{k-1}\cdot 0.97^{11-k}$)<0.05，即有$C_{10}^{0}\cdot 0.03^{0}\cdot 0.97^{10}$ +⋯+$C_{10}^{k-1}\cdot 0.03^{k-1}\cdot 0.97^{11-k}$>0.95，解得 k 的最小值为=2，则技师数=k-1=1。

②.某救援站在长度为 t(/h)的时间内收到救援信号的次数 X 服从 P($\frac{t}{2}$)分布，且与时间的起点无关，求某天下午 1 点到 6 点间至少收到一次救援信号的概率。

P{X≥1}=1-P{X=0}=$1-\frac{\frac{t}{2}^{0}}{0!}\cdot e^{-\frac{t}{2}}=1-\frac{\frac{6-1}{2}^{0}}{0!}\cdot e^{-\frac{6-1}{2}}=1-e^{-\frac{5}{2}}$。

2.3 连续型随机变量及其分布

2.3.1 连续型随机变量的概率密度函数

a.当用函数来表示分布律时，之前用的是概率函数来描述离散型随机变量的分布律，而连续型随机变量的概率分布一般用概率密度函数表示。若存在实数域上的非负可积函数 f(x)，使得随机变量 X 落在区间(a,b]的概率=P{a<X≤b}=$\int_a^b f(x)\cdot dx$，则

f(x)称为 X 的概率密度函数简称为密度函数，并记为 X~f(x)。【古典概型与离散型；几何概型与连续型？】

如果一个可积函数 f(x)满足 $f(x)\geq 0$、$\int_{-\infty}^{+\infty} f(x)\cdot dx=1$，则可将其视为一个连续型 r.v.的密度函数。

*X 的密度函数有时候记为$f_X(x)$，以区别于其他随机变量的密度函数比如随机变量 Y 的密度函数$f_Y(y)$。

*P{X=x}=f(x)=0，f(x)的值并不反映 X 取 x 值的概率，或者说反映，且 X 的各 x 值对应的 f(x)=0；而是：f(x)·dx 反映的是 X 取 x 附近的 x 值的取值概率；另外，虽然 P{X=x}=0，{X=x}并非不可能事件。

*因而对于连续型随机变量，P{a<X≤b}=P{a≤X<b}=P{a<X<b}=P{a≤X≤b}

*由于可改变可列个 f(x)的值，因而对于同一个连续型随机变量，可被不同的密度函数描绘。

*密度函数多为分段函数，对应的区间 G 可能为多个子区间的并集，所以计算概率的积分经常分为分段积分。

2.3.2 常见的连续型随机变量及其分布

1.均匀变量及其分布

$f(x)=\frac{1}{b-a}$，a<x<b；=0，其他，称 X 服从参数为 a,b 的均匀分布，并记为 X~U(a,b)。X 称为均匀变量。

*区间也可为[a,b]、(a,b]、[a,b)。

*(0,1)上的均匀分布在计算机模拟中起重要作用：计算机的伪随机数，是由迭代过程产生的。

2.指数变量及其分布

$f(x)=\lambda\cdot e^{-\lambda x}$，x>0；=0，x≤0，称 X 服从参数为λ的指数分布，并记为 X~E(λ)。X 称为指数变量。

*由于$\int_x^{+\infty}\lambda\cdot e^{-\lambda t}\cdot dt=-e^{-\lambda t}|_x^{+\infty}=e^{-\lambda x}$，因此我们常用这个变下限积分 G(x)=$e^{-\lambda x}$来计算 P{0<X≤x}这个变上限积分所代表的分布函数=$1-e^{-\lambda x}$。

*我们发现，P{X>x}=G(x)具有无记忆性：G(x+y)=G(x)·G(y)，即 P{X>x+y|X>y}=$\frac{P\{X>x+y,\ X>y\}}{P\{X>y\}}=\frac{P\{X>x+y\}}{P\{X>y\}}=\frac{e^{-\lambda(x+y)}}{e^{-\lambda y}}=e^{-\lambda x}$=P{X>x}；而离散型随机变量的分布(密度?)函数中，几何分布也具有无记忆性；那么是不是既具有无记忆性，又是连续型的只能是指数分布呢？——如果成立的话，利用这点，例如蛇等鸟，则只能为指数分布。

确实。由于 G(x+y)=G(x)·G(y)，则 G(x+(m-1)x)=G(x)·G((m-1)x)=G(x)·G(x)·G((m-2)x)=⋯=$G^m(x)$，G(mx)=$G^m(x)$，其中 m 为整数；G($\frac{1}{n}$x+($1-\frac{1}{n}$)x)=G($\frac{1}{n}$x)·G(($1-\frac{1}{n}$)x)= G($\frac{1}{n}$x)·G($\frac{1}{n}$x)·G(($1-\frac{2}{n}$)x)=⋯=$G^n(\frac{1}{n}x)$，G(x)=$G^n(\frac{1}{n}x)$，G($\frac{1}{n}$x)=$G^{\frac{1}{n}}(x)$，其中 n 为整数；综合以上两式，即有 G($\frac{m}{n}$x)=$G^{\frac{m}{n}}(x)$，即 G(Qx)=$G^Q(x)$，其中 Q 为有理数。如果我们再将其拓展为对于不仅对有理数 Q 成立，对于无理数也成立，则我们的函数 G(·)便是定义域在实数域的连续型函数了；另一方面，$\lim\limits_{\triangle x\to 0}\frac{G(x+\triangle x)-G(x)}{\triangle x}=\lim\limits_{\triangle x\to 0}\frac{G(x)G(\triangle x)-G(x)}{\triangle x}=\lim\limits_{\triangle x\to 0}\frac{G(x)[G(\triangle x)-1]}{\triangle x}$，一般地，G(x)大多数地方不为 0，而 G′(x)却在大多数地方为有限值，因此$\lim\limits_{\triangle x\to 0}G(\triangle x)-1$=0，得到$G(0)=1$。

满足这两点的函数，只有指数函数。再在区间(0,+∞)上使用归一化条件$K(x)|_0^{+\infty}=1$，即可知道 K(x)是个负的负指数函数$-a^{-x}$加上一个常数 C。这恰好就是指数分布的密度函数所对应的分布函数，即 K(x)=$1-e^{-\lambda x}$(注：λ的可变和 a 的待定是一回事，因此$-e^{-\lambda x}$和$-a^{-x}$等效)。

****

2′.有关指数分布的一道习题

①.设顾客在某银行窗口等待服务的时间服从=0.2 的指数分布，顾客在窗口等待超过 10min 就离开。

(1).某顾客某天去银行，他未等到服务就离开的概率=$e^{-0.2\cdot 10}$

(2).某顾客一个月去 5 次银行，他五次中至多有一次未等到服务而离开的概率=$C_5^0\cdot(1-e^{-2})^5+C_5^1\cdot e^{-2}\cdot(1-e^{-2})^4$

3.正态变量及其分布

f(x)=$\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$，称 X 服从参数为$\mu,\sigma^2$的正态分布，并记为 X~N($\mu,\sigma^2$)。X 称为正态变量。【这里 N 是 Normal 的意思】

正态变量的密度函数 f(x)有以下性质:

(1).曲线 y=f(x)关于 x=μ对称，表明 P{x<μ}=P{x≥μ}=0.5。

(2).函数 y=f(x)在 x=μ时取最大值$\frac{1}{\sqrt{2\pi}\sigma}$，表面正态变量在 x=$\mu$附近取值的可能性最大。

(3).当σ^2较大时，曲线平坦，表明正态变量的取值比较分散。

*求导可知$\mu\pm\sigma$为 f(x)的两个拐点；f(x)可积不可表。

*温度是正态分布，因为温度源于分子碰撞。

*若 X~N(μ,σ^2)，则 P{a<X≤b}=$\int_a^b \frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\cdot dx=\int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{y^2}{2}}\cdot dy$=P{$\frac{a-\mu}{\sigma}$<Y≤$\frac{b-\mu}{\sigma}$}，其中 Y~N(0,1)。Y 被称为标准正态变量。

*令$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{y^2}{2}}\cdot dy$，则$\Phi$在$x$处的函数值即为标准正态变量 Y 在(-∞,x]内取值的概率。$\Phi(x)$已被制成标准正态分布函数值表。因此 P{$\frac{a-\mu}{\sigma}$<Y≤$\frac{b-\mu}{\sigma}$}=$\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$。

*P{|X − μ|≤σ}=P{-1≤Y≤1}=$\Phi(1)-\Phi(-1)=\Phi(1)-(1-\Phi(1))$=2$\Phi(1)-1$=0.6826

2.4 随机变量的分布函数

2.4.1 r.v.的分布函数

我们之前研究的随机变量有显著差异，表示方式明显不同。而现在我们退一步，仅基于随机变量这个概念，以统一的方式研究各种类型的随机变量。分布函数是这一方式的工具。

一个与随机变量 X 的类型无关，只与 X 所表达的特定事件的概率有关的一元函数 F(x)=P{X≤x}，称为随机变量 X 的分布函数。这里{X≤x}表示 X 的取值不超过 x 的事件，函数 F(x)在 x 处的函数值，即为事件{X≤x}发生的概率。

(1).若 X 是离散型随机变量，则:

F(x)=0，x<x_1；$=p_1$，x_1≤x<x_2；$=p_1+p_2$，x_2≤x<x_3；⋯；

并且对于任意 x，有 P{X=x}=$F(x)-F(x^-)$。

*离散型随机变量的分布函数是右连续、单调不减、阶跃函数。

(2).若 X 是连续型随机变量，则：

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

并且对于任意连续点 x，有 $f(x)=\frac{dF(x)}{dx}$。

*随机变量的分布函数是连续、单调不减函数。

(3).若 X 既不是连续型也不是离散型的随机变量，称之为混合型随机变量。它的分布函数既不是阶跃函数也不是连续函数，仅仅是单调不减的右连续函数。

2.4.2 随机变量的分布函数性质

如果一个函数 F(x)符合 i.$0\le F(x)\le 1$、ii.对于任意实数$x_1<x_2$，都有 $F(x_1)\le F(x_2)$、iii.$F(-\infty)=0$，$F(+\infty)=1$、iv.$F(x^+)=F(x)$，则可将其视为一个 r.v.的分布函数。

与随机变量 X 的类型无关，利用其分布函数 F(x)总可以计算诸如：

$P\{a<X\le b\}=P\{X\le b\}-P\{X\le a\}=F(b)-F(a)$、$P\{a\le X<b\}=P\{a^-<X\le b^-\}=P\{X\le b^-\}-P\{X\le a^-\}=F(b^-)-F(a^-)$、$P\{X=x\}=P\{x^-<X\le x\}=F(x)-F(x^-)$【利用 $P(A-B)=P(A)-P(AB)$，可以得到 $P\{a<X\le b\}=P\{\{X>a\}\cap\{X\le b\}\}=P\{\{X>a\}-\{X>b\}\}=P\{X>a\}-P\{\{X>a\}\cap\{X>b\}\}=P\{X>a\}-P\{X>b\}=P\{X\le b\}-P\{X\le a\}$】

2.5 随机变量的函数的分布

随机变量的函数的分布问题：若已知随机变量 X 的概率分布，求随机变量 Y=g(X)的概率分布。

下面针对 X 是连续型 r.v.或离散型 r.v.时，讨论当 g(x)是连续函数或非连续函数时，Y=g(X)的概率分布。

2.5.1 连续函数的分布 (g(x)是连续函数时)

1.随机变量为离散型时

此时由连续函数 g(x)产生的 Y=g(X)也是离散型随机变量。

具体求解步骤如下：先根据 X 的分布律列出 X 的概率分布表，每一个 P{X=x_i}=p_i竖向对应横向取值不同的 X=x_i，同时也对应着横向取值可能相同的 Y=g(X)=g(x_i)；现将所有相同 g(x_i)所对应的 X=x_i所对应的 P{X=x_i}加起来，作为同一个 Y=g(X)=g(x_i)的概率值 P{Y=g(x_i)}，之后重新列出 Y 的概率分布表即可。

2.随机变量为连续型时

此时由连续函数 g(x)产生的 Y=g(X)也是连续型随机变量。

具体求解步骤如下：

(1).通过 X 的分布函数，求 Y 的分布函数

$F_Y(y)$=P{Y≤y}=P{g(X)≤y}=P{X∈$\cup_i L_i(y)$}=$\sum_i P\{X\in L_i(y)\}$=$\sum_i P\{m_i(y)\le X\le M_i(y)\}$=$\sum_i\int_{m_i(y)}^{M_i(y)} f_X(x)\cdot dx$或=$\sum_i F_X(M_i(y))-F_X(m_i(y))$

(2).求 Y 的密度函数

$f_Y(y)=\frac{dF_Y(y)}{dy}$

*如果随机变量 X 的全部可能取值为区间(a,b)，在之外的$f_X(x)$均=0，且函数 y=g(x)的一阶导数 g′(x)在此区间内恒正或恒负，则可以导出 Y=g(X)的密度函数为：

若 g′(x)>0：$F_Y(y)$=P{Y≤y}=P{g(X)≤y}=P{X≤$g^{-1}(y)$}=$F_X(g^{-1}(y))$，于是$f_Y(y)=\frac{dF_Y(y)}{dy}=\frac{dF_X(g^{-1}(y))}{dy}=(g^{-1}(y))'\cdot f_X(g^{-1}(y))$；

若 g′(x)<0：$F_Y(y)$=P{Y≤y}=P{g(X)≤y}=P{X≥$g^{-1}(y)$}=$1-F_X(g^{-1}(y))$，于是$f_Y(y)=\frac{dF_Y(y)}{dy}=-\frac{dF_X(g^{-1}(y))}{dy}=-(g^{-1}(y))'\cdot f_X(g^{-1}(y))$；

由于$f_Y(y)$必须为正，且$f_X(g^{-1}(y))$一定是>0 的，所以我们将最终结果写为：$f_Y(y)=|(g^{-1}(y))'|\cdot f_X(g^{-1}(y))$，min{g(a),g(b)}<y<max{g(a),g(b)}；=0，其他。【区间也可为[a,b]、(a,b]、[a,b)】

*若$F_X(x)$为严格单调连续函数，证明 $Y=g(X)=F_X(X)$在[0,1]上服从均匀分布：

由于$F_X(x)$不仅是 g(x)，同时也是 X 的分布函数，因此这里 $g'(x)=F'_X(x)>0$，不会<0。

那么：$F_Y(y)=P\{Y\le y\}=P\{F_X(X)\le y\}=P\{X\le F_X^{-1}(y)\}=F_X\left(F_X^{-1}(y)\right)=y$，于是 $f_Y(y)=\frac{dF_Y(y)}{dy}=\frac{dy}{dy}=1$；

****

1.对 $Y=X^2$，$F_Y(y)=P\{Y\le y\}=P\{X^2\le y\}=P\{-\sqrt{y}\le X\le\sqrt{y}\}=F_X(\sqrt{y})-F_X(-\sqrt{y})$，而 $Y=X^2\ge0$，因此$f_Y(y)=\frac{1}{2\sqrt{y}}f_X(\sqrt{y})+\frac{1}{2\sqrt{y}}f_X(-\sqrt{y})$，y>0；0，y≤0。对于$f_X(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$，将其代入即有$f_Y(y)=\frac{1}{\sqrt{2\pi}}\cdot y^{-0.5}\cdot e^{-\frac{y}{2}}$，y>0；0，y≤0。【自由度为 1 的卡方分布】

2.对 $Y=\sin X$，$F_Y(y)=P\{Y\le y\}=P\{\sin X\le y\}=P\{0\le X\le \arcsin y\}+P\{\pi-\arcsin y\le X\le\pi\}$

2.5.2 非连续函数的分布 (g(x)不是连续函数时)

1.随机变量为离散型时

此时由非连续函数 g(x)产生的 Y=g(X)也是离散型随机变量。

2.随机变量为连续型时

此时根据 g(x)的具体表达式，由非连续函数 g(x)产生的 Y=g(X)可能为离散型、连续型、混合型随机变量。

第三章 二维随机变量及其分布

3.1 多维随机变量的概念

有些试验结果需要同时用两个或多个随机变量描述。这种用于描述一个试验结果的多个随机变量称为多维随机变量(随机向量)，记作(X,Y,⋯,W)。研究多维随机变量，不仅要研究其中每一个随机变量的性质，更重要的是研究随机变量之间的相关性质。

3.2 二维离散型随机变量及其分布

3.2.1 二维离散型随机变量的联合分布律

1.二维离散型随机变量(X,Y)的概率分布，又叫(X,Y)的分布律，或随机变量 X 与 Y 的联合分布律。

(X,Y)的概率函数的一般形式为：$f(x_i,y_j)=P\{X=x_i,Y=y_j\}=p_{ij}$，i,j=1,2,⋯.【若(x,y)不是(X,Y)的可能取值，则规定 f(x,y)=P{X=x,Y=y}=0】

它表示(X,Y)的所有取值为(x_i,y_j)，且(X,Y)取(x_i,y_j)的概率为p_{ij}。其中，$\{X=x_i,Y=y_j\}$表示事件$\{X=x_i\}$与事件$\{Y=y_j\}$的积事件。【数学中一般无“、”顿号，并且一般数学上的“,”逗号指“交(集)”，这里即是；不过“i,j=1,2,⋯”却相当于“、”顿号，即“并集”、“或”的意思】

2.概率分布表有“条形”和“矩形”两种：“条形”表能准确且直接地描述一个个(X,Y)的取值及其对应的概率分布状况，相当于将(X,Y)看作一个有名有姓(独一无二)的整体变量，因而这和一维随机变量的分布表很像；“矩形”表能在描述 (X,Y)的取值及其概率分布状况的同时，又能清楚地反映出 X 和 Y 各自的取值状况，并间接反映出它们各自的概率分布状况(需人工求和)。

3.离散型二维 r.v.的概率分布律——即 X 与 Y 的联合分布律，具有如下性质：$p_{ij} \geq 0$，$i,j=1,2,\cdots$、$\sum_i \sum_j p_{ij} = 1$。

3.2.2 边缘分布律与条件分布律

1.边缘分布律

二维离散型随机变量(X,Y)的分量 X 和 Y 都是一维离散型随机变量，X 的分布律和 Y 的分布律分别称为(X,Y)关于 X 的和关于 Y 的边缘分布律，常用概率函数和概率分布表(在表格右边从上到下和下边缘从左往右分别填入一个个行方向和列方向的累计求和结果)表示。

由于 P{X=x_i}满足①.P{X=x_i}=P{X=x_i,-∞<Y<+∞}=$\sum_j P\{X = x_i, Y = y_j\}$=$\sum_j p_{ij}$ ≥0，②.$\sum_i P\{X = x_i\}$=$\sum_i \sum_j p_{ij}$=1，这两条一维离散型随机变量分布律的性质，因此 P{X=x_i}=$\sum_j p_{ij}$=$p_{i\cdot}$，$i=1,2,\cdots$，是(X,Y)关于 X 的边缘分布律。同理(X,Y)关于 Y 的边缘分布律为$p_{\cdot j}$。

联合分布律决定边缘分布律，而由边缘分布律一般不能确定联合分布律。比如二维离散型随机向量的概率分布表中，若有 2*4=8 个随机向量，此时却只有 2+4+1=6 个方程(其中的 1 表示“8 个向量所对应的概率之和为 1”这个方程)。

****

1'.帕斯卡分布

射手进行射击，击中概率为 p，射击击中 r 次为止。X 为首次击中时的射击次数，Y 为总共进行的射击次数。{X=m,Y=n}表示 1,2,⋯,m,⋯,n-1,n 这样一个事件，其中 m 次时首次击中，n 次时共击中 r 次，相当于 m~n 之间还有 r-2 次击中，并且 n≥m+r-1。

在不考虑 X 分量，只看 Y 时，可知在最后一次的射中之前，射击了 n-1 次，且共击中了 r-1 次，因此 P{Y=n}=$C_{n-1}^{r-1} \cdot p^r \cdot (1-p)^{k-r}$。r=1 时，为几何分布。

考虑了 X 分量时，我们来算算(X,Y)关于 X 的边缘分布律 P{X=m}=$\sum_{n=m+r-1}^{\infty} P\{X = m, Y = n\}$=$\sum_{n=m+r-1}^{\infty} C_{n-m-1}^{r-2} \cdot p^r \cdot (1-p)^{n-r}$=$p^2 \cdot (1-p)^{m-1} \sum_{n=m+r-1}^{\infty} C_{n-m-1}^{r-2} \cdot p^{r-2} \cdot (1-p)^{n-r-m+1}$=$p^2 \cdot (1-p)^{m-1}$$p$=$p \cdot (1-p)^{m-1}$。【注：这里用到了第一章章末习题中的$\sum_{i=0}^{\infty} C_{i+j}^{j} \cdot p^j (1-p)^i$=$\frac{1}{p}$】

(X,Y)关于 Y 的边缘分布律 $P\{Y=n\}=\sum_{m=1}^{n-r+1}P\{X=m,Y=n\}=\sum_{m=1}^{n-r+1}C_{n-m-1}^{r-2}\cdot p^r\cdot(1-p)^{n-r}=p^r\cdot(1-p)^{k-r}\cdot\sum_{m=n-r+1}^{1}C_{n-m-1}^{r-2}=C_{n-1}^{r-1}\cdot p^r\cdot(1-p)^{k-r}$，这和我们之前得到的结果一样。

2.条件分布律

若$p_{\cdot j}>0$，则对于 $P\{X=x_i|Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}}$，$j=1,2,\cdots$(j 相对固定)，由于$\frac{p_{ij}}{p_{\cdot j}}\geq 0$且$\sum_i\frac{p_{ij}}{p_{\cdot j}}=1$，因而它描述了在事件$\{Y=y_j\}$发生的条件下，随机变量 X 取值的概率分布律，故称之为在$Y=y_j$条件下的，X 的条件分布律。

3.3 二维连续型随机变量及其分布

3.3.1 二维连续型随机变量的联合概率密度

若存在二维实空间R^2中的一个非负可积函数 f(x,y)，使得对于任意二维区域 G，随机变量(X,Y)落在 G 内的概率为 $P\{(X,Y)\in G\}=\iint_G f(x,y)d\sigma$，那么(X,Y)是二维连续型随机变量。

其中的非负函数$f(x,y)$即为描述二维连续型变量(X,Y)取值规律的概率函数，称为(X,Y)的概率密度函数，简称密度函数，并记$(X,Y)\sim f(x,y)$。$f(x,y)$又称为随机变量 X 与 Y 的联合密度函数。

概率密度函数$f(x,y)$有以下性质：①.$f(x,y)\geq 0$，②.$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=1$。

3.3.2 边缘概率密度函数与条件概率密度函数

1.边缘概率密度函数

二维连续型随机变量(X,Y)的分量 X 和 Y 都是一维连续型随机变量，X 和 Y 的概率分布常用概率函数表示。称为(X,Y)关于 X 和关于 Y 的边缘概率密度函数，简称边缘密度函数。

由于①.$P\{a<X\leq b\}=P\{a<X\leq b,-\infty<Y<+\infty\}=\int_a^b dx\int_{-\infty}^{+\infty}f(x,y)dy$，由于其中的$f(x,y)\geq 0$，则$\int_{-\infty}^{+\infty}f(x,y)dy\geq 0$，②.$\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}f(x,y)dy=1$，因此$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$，

是(X,Y)关于 X 的边缘密度函数。同理(X,Y)关于 Y 的边缘密度函数为$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx$。

联合密度函数决定边缘密度函数。

2.条件概率密度函数

若$f_Y(y)>0$，则对于 $P\{a<X\le b|Y=y\}=\lim_{\triangle y\to 0^+} P\{a<X\le b|y-\triangle y<Y\le y\}=\lim_{\triangle y\to 0^+}\frac{P\{a<X\le b,y-\triangle y<Y\le y\}}{P\{y-\triangle y<Y\le y\}}=\lim_{\triangle y\to 0^+}\frac{\int_a^b dx\int_{y-\triangle y}^{y} f(x,t)dt}{\int_{-\infty}^{+\infty} dx\int_{y-\triangle y}^{y} f(x,t)dt}=\frac{\int_a^b f(x,y)dx}{\int_{-\infty}^{+\infty} f(x,y)dx}=\frac{\int_a^b f(x,y)dx}{f_Y(y)}=\int_a^b\frac{f(x,y)}{f_Y(y)}dx$，由于$\frac{f(x,y)}{f_Y(y)}\ge 0$ 且$\int_a^b\frac{f(x,y)}{f_Y(y)}dx=1$，因而$\frac{f(x,y)}{f_Y(y)}$描述了在事件$\{y^+<Y\le y\}$发生的条件下，事件$\{x^+<X\le x\}$发生的概率，故称$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$为在$Y=y$条件下的，X 的条件密度函数。

【可以这么理解，正如$f_Y(y)$中的 Y 是对 Y 所独有的密度函数$f_Y(\cdot)$的标识；$f_{X|Y}(x|y)$中的X|Y既表示对条件密度函数$f_{X|Y}(x|y)$的标识，又是对$Y=?$条件下的，X 的条件密度函数的标识；并且其中$(x|y)$中的y表示的是条件$Y=y$——之所以$f_{X|Y}(x,y)$不被写为$f_Y(x|y)$，一方面是此时的条件密度函数$f_{X|Y}(x|y)$会变为$f_Y(x|y)$，或者被误以为$f_Y(\cdot)$，而与$f_Y(\cdot)$容易混淆；另一方面可能为了说明这对于 X 来说是密度函数，即在$X=x$时$f_{X|Y}(x|y)\cdot 0=\frac{f(x,y)}{f_Y(y)}\cdot 0=0$(取单个点(X,Y)的概率为 0)；但对于 Y 来说，又有点条件的意思，即分母的$f_Y(y)$即使在$Y=y$时乘以 0 后=0(取单条曲线上的(X,Y)的概率为 0)，整个分式也是有意义的(分子分母同时乘以△ y后，分子再乘以一个△ x，则变成两个无穷小的比较)】

3.3.3 常见的二维连续型随机变量

1.二维均匀变量及其分布

$f(x,y)=\frac{1}{A}$，$(x,y)\in G$；$=0$，其他，其中 A 为区域 G 的面积。称(X,Y)服从区域 G 上的二维均匀分布，并记为(X,Y)~U(G)。

2.正态变量及其分布

$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$，其中均$\sigma_1$、$\sigma_2$均>0，$1<\rho<-1$。称(X,Y)服从参数为$\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho$的二维正态分布，并记为$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$。

注意到$[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]=[\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}]^2+(1-\rho^2)\cdot\frac{(x-\mu_1)^2}{\sigma_1^2}$ $=(1-\rho^2)\cdot t^2+(1-\rho^2)\cdot\frac{(x-\mu_1)^2}{\sigma_1^2}$。其中 $t=\frac{\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}}{\sqrt{1-\rho^2}}$，则 $dt=\frac{dy}{\sigma_2\sqrt{1-\rho^2}}$，于是 $dy=\sigma_2\sqrt{1-\rho^2}\cdot dt$。那么$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2(1-\rho^2)}[(1-\rho^2)\cdot t^2+(1-\rho^2)\cdot\frac{(x-\mu_1)^2}{\sigma_1^2}]}\sigma_2\sqrt{1-\rho^2}\cdot dt=\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1}\cdot e^{-\frac{1}{2}[t^2+\frac{(x-\mu_1)^2}{\sigma_1^2}]}\cdot dt=\frac{1}{2\pi\sigma_1}\cdot e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}e^{-\frac{1}{2}t^2}\cdot dt=\frac{\sqrt{2\pi}}{2\pi\sigma_1}\cdot e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}t^2}\cdot dt=\frac{1}{\sqrt{2\pi}\sigma_1}\cdot e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$。

可见，二维正态变量(X,Y)的分量 X 和 Y 均为一维正态变量。且 $X\sim N(\mu_1,\sigma_1^2)$、$X\sim N(\mu_2,\sigma_2^2)$。注意到，对于参数$\rho$的不同取值，联合密度不同，但边缘密度却相同，这个例子也说明边缘密度一般不能决定联合密度。

3.4 二维随机变量的分布函数

3.4.1 二维 r.v.的联合分布函数

F(x,y)=P{X≤x,Y≤x}，称为二维随机变量(X,Y)的分布函数，或称为随机变量 X 和 Y 的联合分布函数。这里{X≤x,Y≤x}表示 X 的取值不超过 x 且 Y 的取值不超过 y 的积事件。而函数 F 在(x,y)的取值即为事件{X≤x,Y≤x}发生的概率。

(1).若(X,Y)是二维离散型随机变量，则：

$F(X,Y)=\sum_{x_i\le x}\sum_{y_i\le y}P\{X=x_i,Y=y_i\}=\sum_{x_i\le x}\sum_{y_i\le y}p_{ij}$；

并且对于任意 x,y，有 $P\{X=x,Y=y\}=F(x,y)-F(x^-,y)-F(x,y^-)+F(x^-,y^-)$。

*二维离散型随机变量的分布函数是分别关于两个变量的右连续、单调不减、阶跃函数。

(2).若(X,Y)是二维连续型随机变量，则：

$F(x)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)\cdot dxdy$；

并且对于任意连续点 x，有 $f(x,y)=\frac{\partial F(x,y)}{\partial x\,\partial y}$。

*二维连续型随机变量的分布函数是关于两个变量的连续、单调不减函数。

3.4.2 二维随机变量的分布函数性质

二维 r.v.的分布函数 F(x,y)具有如下性质：i.$0\le F(x,y)\le 1$、ii.对于任意实数 $x_1<x_2$、$y_1<y_2$，都有 $F(x_1,y)\le F(x_2,y)$，$F(x,y_1)\le F(x,y_2)$、iii.$F(-\infty,y)=0$，$F(x,-\infty)=0$，$F(+\infty,+\infty)=1$、iv.$F(x^+,y)=F(x,y^+)=F(x,y)$。

与随机变量(X,Y)的类型无关，利用其分布函数 F(X,Y)总可以计算诸如：

$P\{x_1<X\le x_2, y_1<Y\le y_2\}=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)$、$P\{x_1\le X<x_2, y_1\le Y<y_2\}=F(x_2^-,y_2^-)-F(x_1^-,y_2^-)-F(x_2^-,y_1^-)+F(x_1^-,y_1^-)$、$P\{X=x,Y=y\}=P\{x^-<X\le x,y^-<Y\le y\}=F(x,y)-F(x^-,y)-F(x,y^-)+F(x^-,y^-)$

3.4.3 边缘分布函数与条件分布函数

1.边缘分布函数

分量 X 和 Y 的分布函数，称为(X,Y)关于 X 和关于 Y 的边缘分布函数.

$P\{X\le x\}=P\{X\le x,Y<+\infty\}=F(x,+\infty)=F_X(x)$，同理关于 Y 的边缘分布函数为 $F_Y(y)=F(+\infty,y)$。联合分布函数决定边缘分布函数。

2.条件分布函数

在 Y=y 条件下 X 的分布函数$F_{Y|X}(y|x)$【像条件密度函数$f_{X|Y}(x|y)$，括号里的元素之间用的不是逗号而是竖线，因此条件分布函数也并不是这样$F_{Y|X}(y,x)$】，以及在 X=x 条件下 Y 的分布函数$F_{X|Y}(x|y)$，称为(X,Y)的条件分布函数。

离散状况下，$F_{X|Y}(x|y)=P\{X\le x|Y=y\}=\frac{P\{X\le x,Y=y\}}{P\{Y=y\}}=\frac{\sum_{x_i\le x}p_{ij}}{p_{\cdot j}}$。

连续状况下，$F_{X|Y}(x|y)=P\{X\le x|Y=y\}=\frac{\int_{-\infty}^{x}f(x,y)\cdot dx}{\int_{-\infty}^{+\infty}f(x,y)\cdot dx}=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}dx=\int_{-\infty}^{x}f_{X|Y}(x|y)\,dx$

3.5 随机变量的独立性

1.两个变量的独立性

两个随机变量 X,Y 相互独立是指这两个变量之间没有任何关系，即关于 X 的任意一个事件均与关于 Y 的任意一个事件都相互独立。所以不论变量类型是离散的还是连续的，总有两个变量独立的充要条件为：$P\{X\le x,Y\le y\}=P\{X\le x\}\cdot P\{Y\le y\}$，即 $F(x,y)=F_X(x)\cdot F_Y(y)$，或者$F_X(x)=\frac{F(x,y)}{F_Y(y)}$；以及更小范围 $P\{X\le x,Y=y\}=P\{X\le x\}\cdot P\{Y=y\}$，即$F_X(x)=P\{X\le x\}=\frac{P\{X\le x,Y=y\}}{P\{Y=y\}}=P\{X\le x|Y=y\}=F_{X|Y}(x|y)$。——可见 X、Y 的独立会导致$\frac{F(x,y)}{F_Y(y)}=F_{X|Y}(x|y)=\frac{\sum_{x_i\le x}p_{ij}}{p_{\cdot j}}=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}\cdot dx$。再小范围地：

离散型随机变量 X,Y 相互独立的充要条件还有：对于每一对下标 i,j，均有：$P\{X=x_i,Y=y_j\}=P\{X=x_i\}\cdot P\{Y=y_j\}$，或者 $P\{X=x_i\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=P\{X=x_i|Y=y_j\}$。

连续型随机变量 X,Y 相互独立的充要条件还有：对于每一对下标 i,j，均有：$f(x,y)=f_X(x)\cdot f_Y(y)$，或者$f_X(x)=\frac{f(x,y)}{f_Y(y)}=f_{X|Y}(x|y)$。——可见 X、Y 的独立会导致$\int_{-\infty}^{x}f_{X|Y}(x|y)\cdot dx=\int_{-\infty}^{x}f_X(x)=F_X(x)=\frac{F(x,y)}{F_Y(y)}=F_{X|Y}(x|y)$的发生，而这正是条件分布函数的定义。

这时，边缘分布便可决定联合分布了。

2.多个变量的独立性

①.多个随机变量相互独立

(1).充要条件：$F(x_1,x_2,\cdots,x_n)=F_{X_1}(x_1)\cdot F_{X_2}(x_2)\cdots F_{X_n}(x_n)$

(2).离散型的额外充要条件：$P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}=P\{X_1=x_1\}\cdot P\{X_2=x_2\}\cdots P\{X_n=x_n\}$

(3).连续型的额外充要条件：$f(x_1,x_2,\cdots,x_n)=f_{X_1}(x_1)\cdot f_{X_2}(x_2)\cdots f_{X_n}(x_n)$

②.两组随机变量相互独立

若$X_1,X_2,\cdots,X_n$与$Y_1,Y_2,\cdots,Y_n$的联合分布函数 $F(x_1,x_2,\cdots,x_n)$=$X_1,X_2,\cdots,X_n$的联合分布函数 $H(\cdots)$与$Y_1,Y_2,\cdots,Y_n$的联合分布函数 $G(\cdots)$的乘积$=H(y_1,y_2,\cdots,y_n)\cdot G(y_1,y_2,\cdots,y_n)$，

则$X_1,X_2,\cdots,X_n$与$Y_1,Y_2,\cdots,Y_n$相互独立；反之，若$X_1,X_2,\cdots,X_n$与$Y_1,Y_2,\cdots,Y_n$相互独立，则随机变量 H=h($y_1,y_2,\cdots,y_n$)和 G=g($y_1,y_2,\cdots,y_n$)也相互独立。

3.6 二维随机变量的函数的分布

1.(X,Y)为离散型时

此时由连续函数 g 产生的 Z=g(X,Y)也是(一维)离散型随机变量。

具体求解步骤如下：先根据(X,Y)的分布律列出(X,Y)的概率分布表[“条形”或“矩形”]，将所有相同 g(x_i,y_i)所对应的(X,Y)=(x_i,y_i)所对应的p_{ij}加起来，作为同一个 Z=g(X,Y)=g(x_i,y_i)的概率值 P{Z=g(x_i,y_i)}，之后重新列出 Z 的概率分布表即可。

****

(1).离散卷积公式及其应用：

特别地，当 Z=g(X,Y)=X+Y 时，且 X，Y 均能够取 0,1,2,…时，P{Z=r}=P{X+Y=r}=$\sum_{i=0}^{r}P\{X=i,Y=r-i\}$，其中 r 可以取 0,1,2,…。若 X、Y 相互独立的话，还可进一步写为 P{Z=r}=$\sum_{i=0}^{r}P\{X=i\}P\{Y=r-i\}$，r=0,1,2,…。

泊松变量的可加性：由于泊松分布满足 X，Y 均能够取 0,1,2,…，且 X、Y 相互独立，则 P{Z=k}=P{X+Y=k}=$\sum_{i=0}^{k}P\{X=i,Y=k-i\}$=$\sum_{i=0}^{k}P\{X=i\}P\{Y=k-i\}$=$\sum_{i=0}^{k}\frac{\lambda_1^i}{i!}\cdot e^{-\lambda_1}\cdot\frac{\lambda_2^{k-i}}{(k-i)!}\cdot e^{-\lambda_2}$=$\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\cdot\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\cdot\lambda_1^i\cdot\lambda_2^{k-i}$=$\frac{(\lambda_1+\lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)}$

二项变量的可加性：$\sum_{i=0}^{r}P\{X=i,Y=r-i\}$=$\sum_{i=0}^{r}C_{n_1}^{i}\cdot p^i\cdot(1-p)^{n_1-i}\cdot C_{n_2}^{r-i}\cdot p^{r-i}\cdot(1-p)^{n_2-r+i}$=$p^r\cdot(1-p)^{n_1+n_2-r}\cdot\sum_{i=0}^{r}C_{n_1}^{i}\cdot C_{n_2}^{r-i}$=$p^r\cdot(1-p)^{n_1+n_2-r}\cdot C_{n_1+n_2}^{r}$。其中的$\sum_{i=0}^{r}C_{n_1}^{i}\cdot C_{n_2}^{r-i}$=$C_{n_1+n_2}^{r}$用到了如下思想：$(1+x)^{m+n}$=$(1+x)^m(1+x)^n$，将其以二项式展开：$\sum_{j=0}^{m+n}C_{m+n}^{j}\cdot x^j$=$\sum_{j_1=0}^{m}C_m^{j_1}\cdot x^{j_1}\cdot\sum_{j_2=0}^{m}C_m^{j_2}\cdot x^{j_2}$，比较左右两边相同次数的 x($x^j$=$x^{j_1+j_2}$)的系数，即可有之。【或者利用超几何分布有关内容】

几何变量的可加性：若 X~G(p)=NB(1,p)，Y~G(p)=NB(1,p)，且 X,Y 独立，则 Z=X+Y~NB(2,p)，证明如下：P{Z=r}=$\sum_{i=0}^{r}P\{X=i\}P\{Y=r-i\}$=$\sum_{i=1}^{r-1}p(1-p)^{i-1}p(1-p)^{(r-i)-1}$=$\sum_{i=1}^{r-1}p^2(1-p)^{r-2}$=$(r-1)p^2(1-p)^{r-2}$=NB(2,p)，其中 NB(r,p)为帕斯卡分布。

(2).Z=max{X,Y}的分布：

P{Z=z}=P{max{X,Y}=z}=P{X=z,Y≤z}+P{Y=z,X<z}，也可 P{Z=z}=P{Z≤z}-P{Z≤z-1}=P{X≤z,Y≤z}+P{X≤z-1,Y≤z-1}。

2.随机变量为连续型时

此时由连续函数 g 产生的 Z=g(X,Y)也是(一维)连续型随机变量。

具体求解步骤如下：

(1).求 Z 的分布函数

$F_Z(z)$=P{Z≤z}=P{g(X,Y)≤z}=$\iint_{g(x,y)\le z} f(x,y)\cdot dxdy=$ $\iint_{g(x(u,v),y(u,v))\le z} f(x(u,v),y(u,v))\cdot(\frac{\partial x}{\partial u}du,\frac{\partial x}{\partial v}dv)\times(\frac{\partial y}{\partial u}du,\frac{\partial y}{\partial v}dv)=$ $\iint_{g(x(u,v),y(u,v))\le z} f(x(u,v),y(u,v))\cdot(\frac{\partial x}{\partial u},\frac{\partial x}{\partial v})\times(\frac{\partial y}{\partial u},\frac{\partial y}{\partial v})dudv=$ $\iint_{g(x(u,v),y(u,v))\le z} f(x(u,v),y(u,v))\cdot\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}dudv=\iint_{g(x(u,v),y(u,v))\le z} f(x(u,v),y(u,v))\cdot$ $J dudv=\iint_{g(x(u,v),y(u,v))\le z} h(u,v)dudv$

类似 f(x,y)地，其中$h(u,v)=J\cdot f(x(u,v),y(u,v))$为 U、V 的联合密度函数 h(u,v)。

(2).求 Z 的密度函数

$f_Z(z)=\frac{dF_Z(z)}{dz}$；我们也可以通过先求出$h(u,v)=J\cdot f(x(u,v),y(u,v))$的方法来做，其中的 u 相当于 z，是未知量，v 是个补充量。——因此，类似 Z=g(X,Y)地，对称地配对一个 V=v(X,Y)，通过他们找到反函数 X=$x(Z,V)$和 Y=$y(Z,V)$，再求出 Z、V 的联合密度函数$h(z,v)=J\cdot f(x(z,v),y(z,v))$，然后再对联合密度函数偏积分，即得边缘密度函数：$h_Z(z)=\int_{-\infty}^{+\infty}h(z,v)dv$，其中的$h_Z(z)$相当于就是$f_Z(z)$。

****

2'.几个常用的二维连续随机变量的函数分布

a.连续卷积公式及其应用(Z=X+Y 的分布)：

$F_Z(z)$=P{Z≤z}=P{g(X,Y)≤z}=P{X+Y≤z}=$\iint_{x+y\le z} f(x,y)\cdot dxdy=$ $\int_{-\infty}^{+\infty}dy\int_{-\infty}^{z-y}f(x,y)\cdot dx$，令 x=u-y，则$\int_{-\infty}^{+\infty}dy\int_{-\infty}^{z}f(u-y,y)\cdot du=\int_{-\infty}^{z}du\int_{-\infty}^{+\infty}f(u-y,y)dy$，于是$f_Z(z)=\int_{-\infty}^{+\infty}f(z-y,y)dy$，同样的操作设$\int_{-\infty}^{+\infty}dx\int_{-\infty}^{z-x}f(x,y)\cdot dy$和令 y=t-x

仍然可以得到$f_Z(z)=\int_{-\infty}^{+\infty} f(x,z-x)dx$。——或者你也可以通过变量代换 y=z-x，代入$f_Z(z)=\int_{-\infty}^{+\infty} f(z-y,y)dy$，即有$f_Z(z)=\int_{+\infty}^{-\infty} f(x,z-x)d(-x)=\int_{-\infty}^{+\infty} f(x,z-x)dx$。

X、Y 独立时，还额外地有：$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)\cdot f_Y(z-x)dx$和$f_Z(z)=\int_{-\infty}^{+\infty} f_X(z-y)\cdot f_Y(y)dy$。

(1).两个服从相同的均匀分布的相互独立的变量 X,Y 的函数 Z=X+Y 的分布：

设 X,Y 均服从 f(x)=1，0≤x≤1；0，else，则$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)\cdot f_Y(z-x)dx$，0≤x≤1 且 0≤z-x≤1=$\int_0^z 1\cdot dx$，0≤z≤1；$\int_{z-1}^1 1\cdot dx$，1≤z≤2=z，0≤z≤1；2-z，1≤z≤2。

(2).已知(X,Y)~f(x,y)=24xy，x^2≤y≤x；0，else，求 Z=X+Y 的密度函数：

$f_Z(z)=\int_{-\infty}^{+\infty} f(x,z-x)dx$，$x^2$≤z-x≤x=$\int_{\frac{z}{2}}^{\sqrt{z+\frac{1}{4}}-\frac{1}{2}} 24x(z-x)\cdot dx$，$x^2$+x≤z≤2x=$\int_{\frac{z}{2}}^{\sqrt{z+\frac{1}{4}}-\frac{1}{2}} 24x(z-x)\cdot dx$，z>0。

(3).已知(X,Y)~f(x,y)=Cxy，0≤x≤y≤1；0，else，求 Z=X+Y 的密度函数：

$f_Z(z)=\int_{-\infty}^{+\infty} f(x,z-x)dx$，0≤x≤z-x≤1=$\int_{-\infty}^{+\infty} Cx(z-x)dx$，2x≤z≤x+1 且 x>0=$\int_0^{\frac{z}{2}} \frac{1}{8}x(z-x)dx$，0≤z≤1；$\int_{z-1}^{\frac{z}{2}} \frac{1}{8}x(z-x)dx$，1≤z≤2。

(4).已知 X~U(0,1)均匀分布，Y~E(1)指数分布，且独立，Z=X+Y 的密度函数：

由题$f_X(x)$=1，0≤x≤1；0，else；$f_Y Y=e^{-y}$，y>0；0，else。则$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)\cdot f_Y(z-x)dx$，0≤x≤1 且 z-x>0=$\int_0^z e^{-z+x}\cdot dx$，0≤z≤1；$\int_0^1 e^{-z+x}\cdot dx$，z>1=$e^{-z}(e^z-1)$，0≤z≤1；$e^{-z}(e-1)$，z>1=$1-e^{-z}$，0≤z≤1；$e^{1-z}-e^{-z}$，z>1。

(5).已知 X,Y,Z,iid~U(0,1)，求 W=X+Y+Z 的密度函数：【iid：independently and identically distribute，即相互独立且有相同的分布】

先算(X+Y)的，再算 W=(X+Y)+Z 的。

b.类似 Z=X+Y 的分布：$f_Z(z)=\int_{-\infty}^{+\infty} f(z-y,y)dy=\int_{-\infty}^{+\infty} f(x,z-x)dx$。

Z=X-Y 的分布：$f_Z(z)=\int_{-\infty}^{+\infty} f(z+y,y)dy=\int_{-\infty}^{+\infty} f(x,x-z)dx$。

Z=$\frac{X+Y}{2}$的分布：$f_Z(z)=2\int_{-\infty}^{+\infty} f(2z-y,y)dy=2\int_{-\infty}^{+\infty} f(x,2z-x)dx$。

$Z=X^2+Y^2$的分布：$f_Z(z)=\frac{1}{2}e^{-\frac{z}{2}}$，$z>0$，其中 X,Y 独立，且均服从 N(0,1)。【$F_Z(z)=1-e^{-\frac{z}{2}}$，$z>0$】——这里可以和我们之前在 2.5.1 连续函数的分布给出的 $Y=X^2$的$f_Y(y)=\frac{1}{\sqrt{2\pi}}\cdot y^{-0.5}\cdot e^{-\frac{y}{2}}$，y>0 相比较。【自由度为 2 的卡方分布】

$Z=\sqrt{X^2+Y^2}$的分布：$f_Z(z)=z\cdot e^{-\frac{z^2}{2}}$，$z>0$，其中 X,Y 独立，且均服从 N(0,1)。【$F_Z(z)=1-e^{-\frac{z^2}{2}}$，$z>0$】

c.Z=max{X,Y}的分布：

$F_Z(z)$=P{Z≤z}=P{g(X,Y)≤z}=P{max(X,Y)≤z}=P{X≤z,Y≤z}=F(z,z)，于是 $f_Z(z)=\frac{dF_Z(z)}{dz}=\frac{dF(z,z)}{dz}$。

d.Z=min{X,Y}的分布：

$F_Z(z)$=P{Z≤z}=P{g(X,Y)≤z}=P{min(X,Y)≤z}=1-P{X>z,Y>z}=1-[1-F(+∞,z)-F(z,+∞)+F(z,z)]=F(+∞,z)+F(z,+∞)-F(z,z)=$F_Y(z)+F_X(z)$-F(z,z)，于是 $f_Z(z)=\frac{dF_Z(z)}{dz}=\frac{F_Y(z)+F_X(z)-F(z,z)}{dz}=f_Y(z)+f_X(z)-\frac{dF(z,z)}{dz}$。

e.Γ分布：

α,β>0，$f(x)=\frac{1}{\beta^\alpha\Gamma(\alpha)}\cdot x^{\alpha-1}\cdot e^{-\frac{x}{\beta}}$，x>0；0，else，记为 X~Γ(α,β)。其中，$\Gamma(\alpha)$有如下性质：$\Gamma(\alpha+1)=\int_0^{+\infty}t^\alpha\cdot e^{-t}dt=-\int_0^{+\infty}t^\alpha\cdot d(e^{-t})=t^\alpha\cdot e^{-t}|_0^{+\infty}-\int_0^{+\infty}e^{-t}\cdot d(t^\alpha)=0-\alpha\int_0^{+\infty}e^{-t}\cdot t^{\alpha-1}dt=\alpha\Gamma(\alpha)$，因此$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)=\alpha(\alpha-1)\Gamma(\alpha-1)=\cdots=\alpha!\cdot\Gamma(1)=\alpha!$。——可以利用此处的知识得到，若$X_1\sim\Gamma(\alpha_1,\beta)$、$X_2\sim\Gamma(\alpha_2,\beta)$，且它们互相独立，则$X_1+X_2\sim\Gamma(\alpha_1+\alpha_2,\beta)$。

第四章 随机变量的数字特征

有时不容易确定随机变量的概率分布，或者并不需要完全知道随机变量的概率分布，只需要知道它的一些特征就够了。平均值、偏离程度、相互关联程度，就是随机变量的一些特征，是由随机变量的分布所决定的常数，我们称为随机变量的数字特征。

4.1 数学期望

4.1.1 离散型随机变量的数学期望

1.设离散型随机变量 X 的分布律为 P{X=x_i}=p_i，i=1,2,...，若级数$\sum_i x_i p_i$绝对收敛，或者说$\sum_i |x_i| p_i$收敛，则该级数$\sum_i x_i p_i$称为 X 的数学期望，简称期望或均值，是变量取值的平均值，记为 E(X)。

绝对收敛保证，随机变量的数学期望是唯一的：这是因为，正项级数可随意改变求和次序，但不会改变求和结果。

(1).0-1 分布的期望为参数 p。

(2).X~P(λ)，泊松分布的期望为参数λ。

E(X)=$\sum_{k=0}^{\infty} k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty} k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda}=\lambda\sum_{k-1=0}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\cdot e^{-\lambda}=\lambda$。

(3).X~G(λ)，几何分布的期望为$\frac{1}{p}$。

E(X)=$\sum_{k=0}^{\infty} k\cdot(1-p)^{k-1}\cdot p=p\sum_{k=1}^{\infty} k\cdot(1-p)^{k-1}=p\sum_{k=0}^{\infty}\left[(1-p)^k\right]'=p\left[\sum_{k=0}^{\infty}(1-p)^k\right]'=p\cdot\left[\frac{1}{p}\right]'=p\cdot\frac{d(\frac{1}{p})}{d(1-p)}=p\cdot\frac{1}{p^2}=\frac{1}{p}$。【注：这里的$\left[(1-p)^k\right]'$表示$\frac{d(1-p)^k}{d(1-p)}$】

这和我们之前在第一章章末习题中给出的问题和结果有点像，解决方法却要简单很多。因此这个方法应该可以应用于之，并且那里的方法在这里应该是标准做法。

a.设离散型随机变量 X 的分布律为 P{X=x_i}=p_i，i=1,2,...，若级数$\sum_i g(x_i)p_i$绝对收敛，则该级数$\sum_i g(x_i)p_i$为 Y=g(X)的数学期望 E(Y)。(g 是连续函数)

b.设二维离散型随机变量(X,Y)的分布律为 P{X=x_i,Y=x_j}=p_{ij}，i,j=1,2,...，若级数$\sum_i\sum_j g(x_i,y_j)p_{ij}$绝对收敛，则该级数$\sum_i\sum_j g(x_i,y_j)p_{ij}$为 Z=g(X,Y)的数学期望 E(Z)。(g 是连续函数)

这个定理的意义在于：我们不必求出随机变量的函数的分布(其密度函数)，就能求出随机变量的函数的期望。

****

1'.设俩赌徒，A 者以 p 的概率赢 B 者 b 元钱，以 1-p 的概率输 a 元给 B，为了公平起见，a、b 得满足什么关系?

设 E(X)为 A 所赢的 money 的数学期望，则 E(X)=bp-a(1-p)；同理 B 所赢的 money 的数学期望 E(Y)=a(1-p)-bp。二者应该相等，于是有$\frac{a}{b}=\frac{p}{1-p}$。

4.1.2 连续型随机变量的数学期望

1.设连续型随机变量 X 的密度函数为 f(x)，若积分$\int_{-\infty}^{+\infty} x\cdot f(x)dx$绝对收敛，则该积分值称为 X 的数学期望 E(X)。

(1).X~U(a,b)，均匀分布的期望为$\frac{a+b}{2}$。

$E(X)=\int_a^b x\cdot\frac{1}{b-a}\cdot dx=\frac{a+b}{2}$。

(2).X~E(λ)，指数分布的期望为$\frac{1}{\lambda}$。

$E(X)=\int_0^{+\infty} x\cdot\lambda e^{-\lambda x}\cdot dx=-\int_0^{+\infty} x\cdot d(e^{-\lambda x})=-x\cdot e^{-\lambda x}|_0^{+\infty}+\int_0^{+\infty} e^{-\lambda x}\cdot dx=\frac{1}{-\lambda}\int_0^{+\infty} e^{-\lambda x}\cdot d(-\lambda x)=\frac{1}{\lambda}$。

(3).X~N(0,1)，标准正态分布的期望为 0。(由对称性可知其平均值为 0)

$E(X)=\int_{-\infty}^{+\infty} x\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}\cdot dx=0$。

(4).X~$\chi^2(n)$，卡方分布的期望为n。(利用$\Gamma(\frac{n}{2}+1)=\frac{n}{2}\cdot\Gamma(\frac{n}{2})$所对应的$\Gamma(\frac{n}{2})=\frac{2}{n}\cdot\Gamma(\frac{n}{2}+1)$)

$E(X)=\int_0^{\infty} x\cdot\frac{1}{2^{\frac{n}{2}}\cdot\Gamma(\frac{n}{2})}\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}\cdot dx=\int_0^{\infty} x\cdot\frac{n}{2^{\frac{n}{2}+1}\cdot\Gamma(\frac{n}{2}+1)}\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}\cdot dx=n\int_0^{\infty}\frac{1}{2^{\frac{n}{2}+1}\cdot\Gamma(\frac{n}{2}+1)}\cdot x^{\frac{n}{2}}\cdot e^{-\frac{x}{2}}\cdot dx=n$。

a.设连续型随机变量 X 的密度函数为 f(x)，若积分$\int_{-\infty}^{+\infty} g(x)\cdot f(x)dx$绝对收敛，则该积分为 Y=g(X)的数学期望 E(Y)。(g 是连续函数)

$E(Y)=\int_{-\infty}^{+\infty} f_Y(y)dy=\int_{-\infty}^{+\infty} f_X(g^{-1}(y))|(g^{-1}(y))'|\cdot dy=\int_{-\infty}^{+\infty} g(x)\cdot f(x)dx$。

b.设二维连续型随机变量(X,Y)的密度函数为 f(x,y)，若积分$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\cdot f(x,y)dxdy$绝对收敛，则该积分为 Z=g(X,Y)的数学期望 E(Z)。(g 是连续函数)

4.1.3 数学期望的性质

1.数学期望的性质

(1).E(c)=c，c 为常数。

(2).E(cX)=cE(X)，c 为常数。

(3).E(X+Y)=E(X)+E(Y)。

E(X+Y)=$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)\cdot f(x,y)dxdy$=$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\cdot f(x,y)dxdy$+$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y\cdot f(x,y)dxdy$=$\int_{-\infty}^{+\infty}x\cdot dx\int_{-\infty}^{+\infty}f(x,y)dy$+$\int_{-\infty}^{+\infty}y\cdot dy\int_{-\infty}^{+\infty}f(x,y)dx$=$\int_{-\infty}^{+\infty}xf_X(x)\cdot dx$+$\int_{-\infty}^{+\infty}yf_Y(y)\cdot dy$=E(X)+E(Y)。

推论：E(c_1X_1+c_2X_2+⋯+c_nX_n)=c_1E(X_1)+c_2E(X_2)+⋯+c_nE(X_n)。

(4).若 X,Y 相互独立，则 E(XY)=E(X)·E(Y)。【若先有它 E(XY)=E(X)·E(Y)，X,Y 不一定独立：不一定有 P(XY)=P(X)·P(Y)】

E(XY)=$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(xy)\cdot f(x,y)dxdy$=$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf_X(x)yf_Y(y)dxdy$=$\int_{-\infty}^{+\infty}xf_X(x)dx\int_{-\infty}^{+\infty}yf_Y(y)dy$=E(X)·E(Y)。

推论：E($X_1X_2\cdots X_n$)=E(X_1)·E(X_2)⋯E(X_n)。

2.利用数学期望的性质求数学期望

(1).X~N(μ,σ^2)的期望：E(X)=μ。

$\frac{X-\mu}{\sigma}$~N(0,1)，E($\frac{X-\mu}{\sigma}$)=$\frac{1}{\sigma}$E(X)-E($\frac{\mu}{\sigma}$)=$\frac{1}{\sigma}$E(X)$-\frac{\mu}{\sigma}$=0，于是 E(X)=μ。

(2).X~B(n,p)的期望：E(X)=np。

由二项分布的定义，随机变量 X 是 n 重伯努利试验中事件 A 发生的次数。对于每一个(第 i 个)一重伯努利试验，A 发生次数X_i满足 0-1 分布：X_i=1，A 在第 i 次试验中以概率 p 发生；0，A 在第 i 次试验中以概率 1-p 不发生。

于是 X=$\sum_{i=1}^{n}X_i$，于是 E(X)=E($\sum_{i=1}^{n}X_i$)= $\sum_{i=1}^{n}E(X_i)$=np。

(3).若 X 服从超几何分布，则 E(X)=$n\cdot\frac{M}{N}$。

由超几何分布的定义，从有 M 件次品的 N 件产品中抽取 n 件，X 为抽取的次品数。同样，对于每一次单独抽取，抽出的次品数X_i满足 0-1 分布：$X_i=1$，抽出的第 i 件是次品以概率 P{X_i=1}发生；0，抽出的第 i 件不是次品以概率 1-P{X_i=1}发生。即使是无放回抽取，也有 P{X_i=1}=$\frac{M}{N}$。

则 $X=\sum_{i=1}^{n}X_i$，于是 E(X)=E($\sum_{i=1}^{n}X_i$)= $\sum_{i=1}^{n}E(X_i)=n\cdot\frac{M}{N}$。

****

3.一些利用数学期望的性质求数学期望的题目

1.1,2,⋯,n 任意排成一列，若 k 出现在第 k 个位置上，称为一个巧合，求巧合个数 X 的数学期望。

对于数字 k，其巧合个数$X_k=1$，数字 k 出现在第 k 位；0，else。其中 P{X_k=1}=$\frac{(n-1)!}{n!}=\frac{1}{n}$。于是 $X=\sum_{i=1}^{k}X_i$，于是 E(X)=E($\sum_{i=1}^{k}X_i$)=$\sum_{i=1}^{k}E(X_i)=n\cdot\frac{1}{n}$=1。

2.2 元买一注彩票(在 0~999 号中选一个三位数)，其中只有一注有奖，中奖金额 1000 元，问 k 注彩票的平均收益 E(X)?

设 X 为购买 k 注彩票的收益，X_i为购买一注(第 i 注)彩票的收益，由于即使相当于不放回抽取，对于任何先后任何一次抽取，也恒有 P{X_i=998}=$\frac{1}{1000}$，则 E(X_i)=$998\cdot\frac{1}{1000}-2\cdot\frac{999}{1000}$=-1，则 E(X)=E($\sum_{i=1}^{k}X_i$)=$\sum_{i=1}^{k}E(X_i)=k\cdot-1=-k$。

4.2 方差和标准差

4.2.1 方差

若 E[$(X-E(X))^2$]存在，则称 D(X)=E[$(X-E(X))^2$]为 X 的方差。实际中常用与 X 有相同量纲的量$\sqrt{D(X)}$，称为标准差 or 标方差、均方差、根方差。方差是 X 的取值相对于其均值的分散程度的度量值。

根据期望的性质，可以推导出方差的计算公式：D(X)=E[$X^2-2XE(X)+E^2(X)$] =$E(X^2)-2E^2(X)+E^2(X)$=$E(X^2)-E^2(X)$。

1.离散型随机变量的方差

(1).0-1 分布的方差为 p(1-p)。

$E(X^2)=1^2\cdot p+0^2\cdot(1-p)=p$，则 $D(X)=E(X^2)-E^2(X)=p-p^2=p(1-p)$。

(2).$X\sim P(\lambda)$，泊松分布的方差为参数λ。

$E(X^2)=\sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty}k^2\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty}k\cdot\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty}(k-1)\cdot\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda}+\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda}=\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}\cdot e^{-\lambda}+\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\cdot e^{-\lambda}=\lambda^2+\lambda$。

于是 $D(X)=E(X^2)-E^2(X)=\lambda^2+\lambda-\lambda^2=\lambda$。

(3).$X\sim G(\lambda)$，几何分布的方差为$\frac{1-p}{p^2}$。

$E(X^2)=\sum_{k=0}^{\infty}k^2\cdot(1-p)^{k-1}\cdot p=p\sum_{k=1}^{\infty}k^2\cdot(1-p)^{k-1}=p\sum_{k=0}^{\infty}(1-p)[(1-p)^k]''+p\sum_{k=1}^{\infty}k\cdot(1-p)^{k-1}=p(1-p)\left[\sum_{k=0}^{\infty}(1-p)^k\right]''+\frac{1}{p}=p(1-p)\cdot\left[\frac{1}{p}\right]''+\frac{1}{p}=p(1-p)\cdot\frac{d(\frac{1}{p^2})}{d(1-p)}+\frac{1}{p}=p(1-p)\cdot2\frac{1}{p^3}+\frac{1}{p}=2(1-p)\frac{1}{p^2}+\frac{1}{p}=\frac{2}{p^2}-\frac{1}{p}$。

于是 $D(X)=E(X^2)-E^2(X)=\frac{2}{p^2}-\frac{1}{p}-\frac{1}{p^2}=\frac{1}{p^2}-\frac{1}{p}=\frac{1-p}{p^2}$。

2.连续型随机变量的方差

(1).$X\sim U(a,b)$，均匀分布的方差为$\frac{(b-a)^2}{12}$。

$E(X^2)=\int_a^b x^2\cdot\frac{1}{b-a}\cdot dx=\frac{a^2+ab+b^2}{3}$。则 $D(X)=E(X^2)-E^2(X)=\frac{a^2+ab+b^2}{3}-(\frac{a+b}{2})^2=\frac{(b-a)^2}{12}$。

(2).$X\sim E(\lambda)$，指数分布的方差为$\frac{1}{\lambda^2}$。

$E(X^2)=\int_0^{+\infty}x^2\cdot\lambda e^{-\lambda x}\cdot dx=-\int_0^{+\infty}x^2\cdot d(e^{-\lambda x})=-x^2\cdot e^{-\lambda x}|_0^{+\infty}+2\int_0^{+\infty}x\cdot e^{-\lambda x}\cdot dx=\frac{1}{\lambda}2\int_0^{+\infty}x\cdot\lambda e^{-\lambda x}\cdot dx=\frac{2}{\lambda}\cdot\frac{1}{\lambda}=\frac{2}{\lambda^2}$。

于是 $D(X)=E(X^2)-E^2(X)=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$。

(3).$X\sim N(0,1)$，标准正态分布的方差为 1。

$E(X^2)=\int_{-\infty}^{+\infty}x^2\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}\cdot dx=-\int_{-\infty}^{+\infty}x\cdot\frac{1}{\sqrt{2\pi}}\cdot d(e^{-\frac{x^2}{2}})=-x\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}\cdot dx=0+1=1$。

4.2.2 方差的性质

1.方程的性质

(1).D(c)=0，c 为常数。

(2).D(cX)=c^2D(X)，c 为常数。

D(cX)=$E(c^2X^2)-E^2(cX)$=c^2D(X)

(3).D(X+Y)=D(X)+D(Y)+2E[(X-E(X))(Y-E(Y))]。

D(X+Y)=$E((X+Y)^2)-E^2(X+Y)$=$E(X^2+Y^2+2XY)-(E(X)+E(Y))^2$= D(X)+D(Y)+2E(XY)-2E(X)E(Y)=D(X)+D(Y)+2E[(X-E(X))(Y-E(Y))]。其中 E[(X-E(X))(Y-E(Y))]=E(XY)-E(X)E(Y)=cov(X,Y)。【用原始的定义推更好：D(X+Y)=E[$(X+Y-E(X+Y))^2$]=E[$(X-E(X)+Y-E(Y))^2$]更好】

特别地：1.当 Y=c，c 为常数时，D(X+c)=D(X)。

2.当 X,Y 相互独立时，D(X+Y)=D(X)+D(Y)，这是因为 E[(X-E(X))(Y-E(Y))] =E(XY)-E(X)E(Y)=0。

推论：当$X_1X_2\cdots X_n$均相互独立时，D(c_1X_1+c_2X_2+⋯+c_nX_n)= c_1^2D(X_1)+c_2^2D(X_2)+⋯+c_n^2D(X_n)。【之前这里没有条件限制，现在有：“得相互独立”的条件限制。】

更一般地有：D($\sum_{i=1}^n X_i$)=$\sum_{i=1}^n D(X_i)$+2$\sum_i\sum_{j(i<j)}cov(X_i,X_j)$。

(4).D(X)=0 的充要条件为存在常数 c，使得 P{X=c}=1，且 c=E(x)。【不需要说 P{X≠c}=0，正如我不需要在此加这句话】

证明用到切比雪夫不等式：0≤P{|X-E(X)|≥ε}≤$\frac{D(X)}{\varepsilon^2}$=0，因此 P{|X-E(X)|≥ε}=0，因此 P{|X-E(X)|<ε}≥1−$\frac{D(X)}{\varepsilon^2}$=1。又它不能>1，所以 P{|X-E(X)|<ε}=1。又因ε是任意小的，因此 P{|X-E(X)|=0}=1，即 P{X=E(X)}=1。——且 P{|X-E(X)|>0}=0，即 P{X≠E(X)}=0。

2.利用方差的性质求方差

(1).X~N(μ,σ^2)的方差：D(X)=σ^2。

$\frac{X-\mu}{\sigma}$~N(0,1)，D($\frac{X-\mu}{\sigma}$)=D($\frac{X}{\sigma}$)=$\frac{1}{\sigma^2}$D(X)=1，于是 D(X)=σ^2。

(2).X~B(n,p)的方差：D(X)=np(1-p)。

由二项分布的定义，随机变量 X 是 n 重伯努利试验中事件 A 发生的次数。对于每一个(第 i 个)一重伯努利试验，A 发生次数X_i满足 0-1 分布：X_i=1，A 在第 i 次试验中以概率 p 发生；0，A 在第 i 次试验中以概率 1-p 不发生。

于是 $X=\sum_{i=1}^{n}X_i$，于是 D(X)=D($\sum_{i=1}^{n}X_i$)=$\sum_{i=1}^{n}D(X_i)$=np(1-p)。

(3).若 X 服从超几何分布，则 E(X)=$n\cdot\frac{M}{N}$。

由超几何分布的定义，从有 M 件次品的 N 件产品中抽取 n 件，X 为抽取的次品数。同样，对于每一次单独抽取，抽出的次品数X_i满足 0-1 分布：X_i=1，抽出的第 i 件是次品以概率 P{X_i=1}发生；0，抽出的第 i 件不是次品以概率 1-P{X_i=1}发生。即使是无放回抽取，也有 P{X_i=1}=$\frac{M}{N}$。

则 $X=\sum_{i=1}^{n}X_i$，于是 E(X)=E($\sum_{i=1}^{n}X_i$)= $\sum_{i=1}^{n}E(X_i)=n\cdot\frac{M}{N}$。

(4).若 D(X)≠0，则 X*=$\frac{X-E(X)}{\sqrt{D(X)}}$称为 X 的标准化随机变量。因其 E(X*)=0、D(X*)=1。比如当 X~B(n,p)时，X*=$\frac{X-np}{\sqrt{np(1-p)}}$是 X 的标准化随机变量；当 X~N($\mu,\sigma^2$)时，X*=$\frac{X-\mu}{\sigma}$是 X 的标准化随机变量，且 X*~N(0,1)。

4.2.3 切比雪夫不等式

P{|X-E(X)|≥ε}=$\int_{|x-E(x)|\ge\varepsilon}f(x)dx=\int_{\frac{|x-E(x)|}{\varepsilon}\ge1}1\cdot f(x)dx\le\int_{|x-E(x)|\ge\varepsilon}\left(\frac{|x-E(x)|}{\varepsilon}\right)^2\cdot f(x)dx=\frac{1}{\varepsilon^2}\int_{|x-E(x)|\ge\varepsilon}[x-E(x)]^2\cdot f(x)dx=\frac{E_{|x-E(x)|\ge\varepsilon}([x-E(x)]^2)}{\varepsilon^2}=\frac{D_{|x-E(x)|\ge\varepsilon}(X)}{\varepsilon^2}$。虽然不知道$D_{|x-E(x)|\ge\varepsilon}(X)$(或许不能这么称呼，因为只使用了$|X-E(X)|\ge\varepsilon$区间上的 f(x)进行了积分【不过当$\varepsilon$很小时，$|X-E(X)|\ge\varepsilon$近似于整个定义域】，而 f(x)是对于整个区间而言的，整个区间上的 f(x)积分才为 1)和D(X)的关系(即使D(X)更小)，不过确实有：P{|X-E(X)|≥ε}≤$\frac{D(X)}{\varepsilon^2}$。进而 P{|X-E(X)|<ε}≥1−$\frac{D(X)}{\varepsilon^2}$。这说明，D(X)越小，{|X-E(X)|≥ε}这个事件发生的概率越小。

****

一道奇怪的题：

1.(1).对于密度函数为 f(x)=$\frac{x^m}{m!}\cdot e^{-x}$，x≥0；0，else。【注，这和泊松分布$\frac{\lambda^k}{k!}\cdot e^{-\lambda}$稍有不同，它们恰好自变量和常数对调了。】

$E(X)=\int_0^{+\infty} x\cdot\frac{x^m}{m!}\cdot e^{-x}\cdot dx=-\int_0^{+\infty}\frac{x^{m+1}}{m!}\cdot de^{-x}=-\frac{x^{m+1}}{m!}\cdot e^{-x}|_0^{+\infty}+(m+1)\int_0^{\infty}e^{-x}\cdot\frac{x^m}{m!}dx=0+(m+1)\int_0^{+\infty}f(x)\cdot dx=m+1$

$E(X^2)=\int_0^{+\infty}x^2\cdot\frac{x^m}{m!}\cdot e^{-x}\cdot dx=(m+1)\int_0^{+\infty}x\cdot\frac{x^{m+1}}{(m+1)!}\cdot e^{-x}\cdot dx=(m+1)(m+2)$

于是 $D(X)=(m+1)(m+2)-(m+1)^2=m+1$

(2).接着第一个问，证明 $P\{0<X<2(m+1)\}\geq\frac{m}{m+1}$。

即要证明 $P\{-(m+1)<X-E(X)<m+1\}\geq\frac{m}{m+1}$。根据切比雪夫不等式，那么则有 $P\{|X-E(X)|<m+1\}\geq 1-\frac{D(X)}{\varepsilon^2}=1-\frac{m+1}{(m+1)^2}=\frac{m}{m+1}$。

4.3 协方差和相关系数

4.3.1 协方差及其性质

当 X,Y 是同一个样本空间中的随机变量时，还有必要讨论它们之间的关联程度。协方差在控制问题中大有作为。

1.设(X,Y)为二维随机变量，称 E[(X-E(X))(Y-E(Y))]为与 Y 的协方差，记为 cov(X,Y)。

根据期望的性质，我们之前得到了协方差的简化计算公式：cov(X,Y)=E(XY)-E(X)E(Y)。另外，cov(X,X)=D(X)；当 X,Y 相互独立时，cov(X,Y)=0【期望的性质 4】。

2.协方差的性质

(1).cov(X,Y)=cov(Y,X)

(2).cov(aX,bY)=abcov(Y,X)

cov(aX,bY)=E(aXbY)-E(aX)E(bY)=abE(XY)-abE(X)E(Y)=abcov(Y,X)

(3).cov(X+Y,Z)=cov(X,Z)+cov(Y,Z)

cov(X+Y,Z)=E((X+Y)Z)-E(X+Y)E(Z)=E(XZ)+E(YZ)-E(X)E(Z)-E(Y)E(Z)=cov(X,Z)+cov(Y,Z)

该性质可推广到多个随机变量的情形：$cov(\sum_{i=1}^{n}X_i,Y)=\sum_{i=1}^{n}cov(X_i,Y)$

3.若C_{ij}=cov(X_i,X_j)都存在，则矩阵 **C**=$\begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{pmatrix}$称为($X_1$,$X_2$,⋯,$X_n$)的协方差矩阵。由于$(X_1,X_2,\cdots,X_n)\cdot \mathbf{C}\cdot(X_1,X_2,\cdots,X_n)^T$=f($X_1$,$X_2$,⋯,$X_n$)(即 n 维向量的密度函数)≥0，因此根据定义，**C** 是半正定的。【判断正定与否的方法：1.定义；2.正惯性指数；3.顺序主子式；4.主子式；5.各个特征值>0；6.标准型中 n 个正数；7.存在正交矩阵 P，使得P^TAP=对角阵；8.合同于单位阵 IorE；9.A=UU^T；10.存在上三角阵 P，使得$P^{-1}AP$ =对角阵。】

****

设X_1,X_2,⋯,X_n，iid 独立且服从相同分布，且 D(X_i)=σ^2，求 cov(X_1,$\overline{X}$)和 D($\overline{X}$)。【注：$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，也是随机变量(像$X_i$一样，人家也是大写的)】

cov(X_1,$\overline{X}$)=cov(X_1,$\frac{1}{n}\sum_{i=1}^{n}X_i$)=$\frac{1}{n}$cov($X_1$,$\sum_{i=1}^{n}X_i$)=$\frac{1}{n}\sum_{i=1}^{n}\mathrm{cov}(X_1,X_i)$=$\frac{1}{n}\mathrm{cov}(X_1,X_1)$=$\frac{1}{n}D(X_1)$=$\frac{\sigma^2}{n}$。

正如 D(X+Y)=D(X)+D(Y)+2cov(X,Y)一样，更一般地有 D($\sum_{i=1}^{n}X_i$)=$\sum_{i=1}^{n}D(X_i)$+$2\sum_i\sum_{j(i<j)}\mathrm{cov}(X_i,X_j)$，当$X_i$互相独立时，D($\overline{X}$)=D($\frac{1}{n}\sum_{i=1}^{n}X_i$)=$\frac{1}{n^2}$D($\sum_{i=1}^{n}X_i$)=$\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)$=$\frac{\sigma^2}{n}$。这是一个较重要的结论：cov($X_1$,$\overline{X}$)=D($\overline{X}$)=$\frac{\sigma^2}{n}$。

4.3.2 相关系数及其性质

由于协方差 cov(X,Y)的值会受到 X,Y 的量纲的影响[∵cov(kX,kY)=k^2cov(Y,X)]，因此我们需要把 X,Y 先中心化[比如$X-E(X)$]，再标准化为 X*=$\frac{X-E(X)}{\sqrt{D(X)}}$、Y*=$\frac{Y-E(Y)}{\sqrt{D(Y)}}$。

此时 cov(X*,Y*)=E(X*Y*)-E(X*)E(Y*)=E($\frac{X-E(X)}{\sqrt{D(X)}}\cdot\frac{Y-E(Y)}{\sqrt{D(Y)}}$)-E($\frac{X-E(X)}{\sqrt{D(X)}}$)E($\frac{Y-E(Y)}{\sqrt{D(Y)}}$)=$\frac{1}{\sqrt{D(X)}}\cdot\frac{1}{\sqrt{D(Y)}}$E[(X-E(X))(Y-E(Y))]-0=$\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$，称为 X 与 Y 的相关系数，将其记为$\rho$(X,Y)=cov(X*,Y*)=$\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$，它是个无量纲的数(二维正态分布中的$\rho$就是这么来的)。

若相关系数ρ=0，即 X,Y 不相关，导致$\mathrm{cov}(X,Y)$=0，则 E(XY)-E(X)E(Y)=0，但它并不能推出 X,Y 独立。因此，不相关不能→独立，然而在不考虑分母$\sqrt{D(X)}\sqrt{D(Y)}$中的方差为 0 时，独立却能推出无关。举个不相关但不独立的例子：设 X~U($-\frac{1}{2}$,$-\frac{1}{2}$)，Y=cosX，则 E(X)=0，且 E(XY)=$\int_{-\frac{1}{2}}^{\frac{1}{2}}x\cdot\cos x\cdot dx$=0，则$\mathrm{cov}(X,Y)$=E(XY)-E(X)E(Y)=0，则$\rho$=cov(X*,Y*)=0，即不相关，但是 X 与 Y 有严格的函数关系，即 X 和 Y 不独立。

1.相关系数的性质

(1).$\rho(X,Y)=\rho(Y,X)$

(2).$|\rho(X,Y)|\leq 1$

法一：$0\leq D(Y-bX)=D(Y)+D(-bX)+2cov(Y,-bX)=D(Y)+b^2D(X)-2bcov(Y,X)$，假设其中 $b=\frac{cov(Y,X)}{D(X)}$，则原式$=D(Y)+\frac{cov^2(Y,X)}{D(X)}-2\frac{cov^2(Y,X)}{D(X)}=D(Y)-\frac{cov^2(Y,X)}{D(X)}=D(Y)(1-\rho^2)$，由于 $D(Y)\geq 0$，所以 $1-\rho^2\geq 0$，所以$|\rho|\leq 1$。

法二：引理：对于$\forall t$，$f(t)=E[(Y-tX)^2]=E(X^2)t^2-2E(XY)t+E(Y^2)\geq 0$，因此可得$E^2(XY)\leq E(X^2)E(Y^2)$

令$X_1=X-E(X)$，$Y_1=Y-E(Y)$，则 $E(X_1^2)=D(X)$、$E(Y_1^2)=D(Y)$、$E(X_1Y_1)=cov(X,Y)$，此时$\rho^2=\frac{cov^2(X,Y)}{D(X)D(Y)}=\frac{E^2(X_1Y_1)}{E(X_1^2)E(Y_1^2)}\leq 1$。

(3).$|\rho|=1$ 的充分必要条件为存在常数 a,b，使得 $P\{Y=aX+b\}=1$。另外，当 $a>0$ 时，$\rho(X,Y)=1$；当 $a<0$ 时，$\rho(X,Y)=-1$。

利用引理部分的说法，$\exists t_0=\frac{E(X_1Y_1)}{E(X_1^2)}=\frac{E(Y_1^2)}{E(X_1Y_1)}$ [即$E^2(X_1Y_1)=E(X_1^2)E(Y_1^2)$、$\rho^2=\frac{cov^2(X,Y)}{D(X)D(Y)}=\frac{E^2(X_1Y_1)}{E(X_1^2)E(Y_1^2)}=1$ 时]，使得 $f(t_0)=E[(Y_1-t_0X_1)^2]=0$，于是 $0\leq D(Y_1-t_0X_1)=E[(Y_1-t_0X_1)^2]-E^2(Y_1-t_0X_1)\leq E[(Y_1-t_0X_1)^2]=0$，即存在$t_0$，使得 $D(Y_1-t_0X_1)=0$。

利用方差的第(4).点性质：$D(Y_1-t_0X_1)=0$ 的充要条件为存在常数 c，使得 $P\{Y_1-t_0X_1=c\}=1$，且 $c=E(Y_1-t_0X_1)$。

由于$Y_1-t_0X_1=[Y-E(Y)]-t_0[X-E(X)]=Y-t_0X-[E(Y)+t_0E(X)]$。所以 $c=E(Y_1-t_0X_1)=0$，且$Y_1-t_0X_1=c=0$ 被写为$Y=t_0X+[E(Y)+t_0E(X)]$。令其中 $a=t_0=\frac{E(X_1Y_1)}{E(X_1^2)}=\frac{cov(X,Y)}{D(X)}$，$b=E(Y)+t_0E(X)$，于是 $P\{Y_1-t_0X_1=0\}=1$ 变为了 $P\{Y=t_0X+[E(Y)+t_0E(X)]\}=1$ 进而变为了 $P\{Y=aX+b\}=1$，且其中 $a=\frac{cov(X,Y)}{D(X)}$。

当 $a>0$ 时，$cov(X,Y)>0$，$\rho(X,Y)=\frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}>0$，因此此时$\rho(X,Y)=1$。——或者从另一个角度，$\rho=\frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{cov(X,aX+b)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{cov(X,aX)+cov(X,b)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{acov(X,X)+E(bX)-E(b)E(X)}{\sqrt{D(X)}\sqrt{D(Y)}}=a\frac{cov(X,X)}{\sqrt{D(X)}\sqrt{D(Y)}}$，其中$cov(X,X)=D(X)\geq 0$，因而$\rho$与$a$同号。

****

1.设(X,Y)服从二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求 $cov(X,Y)$，$\rho(X,Y)$。

由之前的知识可知，E(X)=μ_1、E(Y)=μ_2、D(X)=σ_1^2、D(Y)=σ_2^2，因而 cov(X,Y)=E[(X-E(X))(Y-E(Y))]=E[(X-μ_1)(Y-μ_2)]=$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy$ $=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)e^{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]}dxdy$，令其中 u=$\frac{x-\mu_1}{\sigma_1}$，v=$\frac{y-\mu_2}{\sigma_2}$，则原式变为=$\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\cdot\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}uve^{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}}dudv=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\cdot\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}uve^{-\frac{v^2}{2}-\frac{(u-\rho v)^2}{2(1-\rho^2)}}dudv=\int_{-\infty}^{+\infty}ve^{-\frac{v^2}{2}}dv\int_{-\infty}^{+\infty}ue^{-\frac{(u-\rho v)^2}{2(1-\rho^2)}}du=\rho\sigma_1\sigma_2$，于是$\rho$(X,Y)=$\frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2}=\rho$。——可见$\rho$为 X,Y 的相关系数。

从服从 N(μ_1,μ_2,σ_1^2,σ_2^2,ρ)的(X,Y)的密度函数 f(x,y)，以及以前得到的 X,Y 的边缘密度表达式$f_X(x)$、$f_Y(y)$上来看；当ρ≠0 时，恒有 f(x,y)≠$f_X(x)\cdot f_Y(y)$；当ρ=0 时，才有 f(x,y)=$f_X(x)\cdot f_Y(y)$。因此，对于二维正态变量而言，“相互独立”与“不相关”是等价的。

n 维正态变量有以下重要性质：

性质 1：n 维正态变量($X_1,X_2,\cdots,X_n$)的每一个分量X_i都是一维正态变量【偏积分】。反之，若每个X_i都是一维正态变量，且相互独立(即不相关)，则($X_1,X_2,\cdots,X_n$)是 n 维正态变量。

性质 2：n 维随机变量($X_1,X_2,\cdots,X_n$)服从 n 维正态分布的充要条件为$X_1,X_2,\cdots,X_n$的任意线性组合$L_1X_1+L_2X_2+\cdots+L_nX_n$服从一维正态分布。——利用这个性质，可以通过一维正态变量来研究多维正态变量。

——利用以上两个性质，我们还可以得到一个很有意思的结论：根据性质 1：若每个X_i都是一维正态变量，且相互独立(即不相关)，则($X_1,X_2,\cdots,X_n$)是 n 维正态变量；进一步地，由于满足这些条件的X_i所对应的($X_1,X_2,\cdots,X_n$)服从 n 维正态分布，根据性质 2：则其必然推出(其必要条件)$X_1,X_2,\cdots,X_n$的任意线性组合$L_1X_1+L_2X_2+\cdots+L_nX_n$服从一维正态分布。

也就是说，利用 n 维正态变量($X_1,X_2,\cdots,X_n$)这个桥梁，我们便有两组一维正态变量之间的映射：若每个X_i都是一维正态变量，且相互独立(即不相关)，则$X_1,X_2,\cdots,X_n$的任意线性组合 Y=$L_1X_1+L_2X_2+\cdots+L_nX_n$服从一维正态分布。——既然如此，我们只需要求得 Y 的两个数字特征 E(Y)=$\sum_{i=1}^{n}L_iE(X_i)=\sum_{i=1}^{n}L_i\mu_i$和$\sum_{i=1}^{n}L_i^2D(X_i)=\sum_{i=1}^{n}L_i^2\sigma_i^2$，就可以得到 Y 的分布(即密度函数)：N($\sum_{i=1}^{n}L_i\mu_i,\sum_{i=1}^{n}L_i^2\sigma_i^2$)了。——这点性质本该在 3.6 中通过例子来引入(/发现)的，但在这里才能解释。

同样，若$X_1,X_2,\cdots,X_n$的任意线性组合$L_1X_1+L_2X_2+\cdots+L_nX_n$都服从一维正态分布，则 n 维随机变量($X_1,X_2,\cdots,X_n$)服从 n 维正态分布，即($X_1,X_2,\cdots,X_n$)是 n 维正态变量，则 n 维正态变量($X_1,X_2,\cdots,X_n$)的每一个分量$X_i$都是一维正态变量。——即若

$X_1,X_2,\cdots,X_n$的任意线性组合$L_1X_1+L_2X_2+\cdots+L_nX_n$都服从一维正态分布，则每一个$X_i$都是一维正态变量。——于是"$X_1,X_2,\cdots,X_n$的任意线性组合$L_1X_1+L_2X_2+\cdots+L_nX_n$都服从一维正态分布"与"每一个$X_i$都是一维正态变量"互为充要条件。

性质 3：若 n 维随机变量$(X_1,X_2,\cdots,X_n)$服从 n 维正态分布，且$Y_j=L_{1j}X_1+L_{2j}X_2+\cdots+L_{nj}X_n$，则$(Y_1,Y_2,\cdots,Y_m)$服从 m 维正态分布。

性质 4：若 n 维随机变量$(X_1,X_2,\cdots,X_n)$服从 n 维正态分布，则$X_1,X_2,\cdots,X_n$相互独立的充要条件为$X_1,X_2,\cdots,X_n$两两不相关。

4.4 矩

随机变量的幂函数的期望很常见，称 E(X^k)为 X 的 k 阶原点矩，称 E($[X-E(X)]^k$)为 X 的 k 阶中心矩，其中 k 为正整数。E(X)为一阶原点矩，D(X)为二阶中心矩。

对于二维随机变量(X,Y)，称 E(X^kY^L)为(X,Y)的(k,L)阶混合原点矩，称 E($[X-E(X)]^k[Y-E(Y)]^L$)为 X 与 Y 的(k,L)阶混合中心矩，其中 k,L 为正整数。cov(X,Y)为(1,1)阶混合中心矩。

****

设 X~N(0,1)，求 E(X^k)。【这和母函数有关】

E(X^k)=$\int_{-\infty}^{+\infty} x^k\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}\cdot dx$，当 k 为奇数时，被积函数为奇函数，因而 E($X^k$)=0。当 k 为偶数时，

E(X^k)=$-\frac{1}{\sqrt{2\pi}}\cdot\int_{-\infty}^{+\infty} x^{k-1}\cdot e^{-\frac{x^2}{2}}\cdot d(-\frac{x^2}{2})=-\frac{1}{\sqrt{2\pi}}\cdot[x^{k-1}\cdot e^{-\frac{x^2}{2}}|_{-\infty}^{+\infty}-\int_{-\infty}^{+\infty}(k-1)x^{k-2}\cdot e^{-\frac{x^2}{2}}\cdot dx]=(k-1)\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot x^{k-2}\cdot e^{-\frac{x^2}{2}}\cdot dx$=(k-1)·E($X^{k-2}$)=(k-1)·(k-3)·E($X^{k-4}$)=⋯=(k-1)·(k-3)⋯3·E($X^2$)=(k-1)!![D(X)-$E^2$(X)]=(k-1)!![1-0]=(k-1)!!

第五章 大数定理与中心极限定理

5.1 大数定理

在第一章中已经指出：一个随机事件的发生在大量重复试验中会呈现出明显的规律性——即随机事件发生的频率，随着试验次数 n 的增大，逐渐稳定于某个常数。——这种“频率的稳定性”使人们认识到随机事件的概率是客观存在的，并成为概率定义的基础。——伯努利大数定理将以严格的数学形式表达频率的这种稳定性；凡是阐述大量随机现象的平均结果具有稳定性的定理，都称为大数定理。

首先，我们称$X_1,X_2,\cdots,X_n$为一个随机变量序列，称随机变量$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$为随机变量序列$X_1,X_2,\cdots,X_n$的算术平均，它也是个随机变量(∵你看它也是大写的)——这和 X 的数学期望 E(X)=$\sum_i x_i p_i=\sum_i x_i\cdot\frac{1}{n}=\frac{1}{n}\sum_i x_i$，有着完全不同的意义(你看求和符号里的x是不是小写的)，所以在这里，$\overline{X}$并不表示随机变量 X 的数学期望 E(X)or 其平均值，而是随机变量序列$X_1,X_2,\cdots,X_n$的算术平均，是个随机变量。

切比雪夫大数定理：若随机变量序列$X_1,X_2,\cdots,X_n$两两相互独立，方差存在且有共同上界，则对于任意给定小的ε，$\lim_{n\to\infty}P\{|\overline{X}-E(\overline{X})|<\varepsilon\}=1$。——即在以上两个条件下，随着 n 的趋于无穷，$\frac{1}{n}\sum_{i=1}^{n}X_i$与其数学期望相差很小的概率越来越接近于 1——即$X_1,X_2,\cdots,X_n$的算术平均依概率收敛于其算术平均的期望：$\overline{X}\xrightarrow{p}E(\overline{X})$，即$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{p=1}\frac{1}{n}\sum_{i=1}^{n}E(X_i)$。

根据切比雪夫定理，P{|X-E(X)|≥ε}≤$\frac{D(X)}{\varepsilon^2}$，我们使用后者：P{|X-E(X)|<ε}≥$1-\frac{D(X)}{\varepsilon^2}$，将其中的 X 替换为$\overline{X}$，即有 P{|$\overline{X}$-E($\overline{X}$)|<ε}≥$1-\frac{D(\overline{X})}{\varepsilon^2}$。由于随机变量序列$X_1,X_2,\cdots,X_n$两两相互独立，因此$D(\overline{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}D(\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)$，由于随机变量序列$X_1,X_2,\cdots,X_n$的方差有共同上界，即$D(X_i)\le k$，因此$D(\overline{X})=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)\le\frac{nk}{n^2}=\frac{k}{n}$，即有$-D(\overline{X})\ge-\frac{k}{n}$，因此$(1\ge)P\{|\overline{X}-E(\overline{X})|<\varepsilon\}\ge1-\frac{D(\overline{X})}{\varepsilon^2}\ge1-\frac{k}{\varepsilon^2 n}$。那么$\lim_{n\to\infty}P\{|\overline{X}-E(\overline{X})|<\varepsilon\}\ge\lim_{n\to\infty}1-\frac{k}{\varepsilon^2 n}=1$。

依概率收敛：设$\{X_n\}=X_1,X_2,\cdots,X_n$为随机变量序列，若对∀ε>0，都有$\lim_{n\to\infty}P\{|X_n-a|<\varepsilon\}=1$，称$\{X_n\}$依概率p收敛于 a，记为$X_n\xrightarrow{p}a$。【收敛的定义却是：对任意ε>0，存在 N(ε)，使得 n>N 时，$|X_n-a|<\varepsilon$】

1.一个例题

设$X_n \xrightarrow{p} x$，证明$X_nY \xrightarrow{p} xY$，其中X_n与Y独立。

对∀δ>0，总能取到一个较大的数 M>0，使得 P{|Y|≥M}<$\frac{\delta}{2}$；由于$X_n \xrightarrow{p} x$，则根据其定义：对∀ε>0，总可另取到一个 N，当 n>N 时，P{|X_n-x|<$\frac{\varepsilon}{M}$}>1−$\frac{\delta}{2}$，即 P{|X_n-x|≥$\frac{\varepsilon}{M}$}<$\frac{\delta}{2}$；那么 P{|$X_nY - xY$|≥ε}=P{|$X_n - x$||Y|≥ε}=P{|$X_n - x$||Y|≥ε,|Y|≥M}+P{|$X_n - x$||Y|≥ε,|Y|<M}，由于事件的独立性，其中交事件 P{|$X_n - x$||Y|≥ε,|Y|≥M}=积事件 P{|$X_n - x$||Y|≥ε}·P{|Y|≥M}≤P{|Y|≥M}<$\frac{\delta}{2}$，同样的道理 P{|$X_n - x$||Y|≥ε,|Y|<M}=P{|$X_n - x$||Y|≥ε}·P{|Y|<M}≤P{|$X_n - x$|M≥ε}·P{|Y|<M}≤P{|$X_n - x$|≥$\frac{\varepsilon}{M}$}<$\frac{\delta}{2}$，因此，对∀δ>0，当 n>N 时，P{|$X_nY - xY$|≥ε}<$\frac{\delta}{2}+\frac{\delta}{2}$=$\delta$，即$X_nY \xrightarrow{p} xY$。

依概率收敛的随机变量序列的性质：若$X_n \xrightarrow{p} a$，函数 g(x)在 a 处连续，则$g(X_n) \xrightarrow{p} g(a)$；若$X_n \xrightarrow{p} a$，$Y_n \xrightarrow{p} b$，函数 g(x,y)在(a,b)处连续，则$g(X_n, Y_n) \xrightarrow{p} g(a,b)$。

由 g(x,y)在(a,b)处连续的定义，对∀ε>0，存在δ>0，当$\sqrt{(x-a)^2+(y-b)^2}<\delta$时，或者说|x-a|+|y-b|<$\delta$时，总有|g(x,y)-g(a,b)|<$\varepsilon$。——于是{|g(x,y)-g(a,b)|≥$\varepsilon$}⊂|x-a|+|y-b|≥$\delta$ ⊂{|x-a|≥$\frac{\delta}{2}$}∪{|y-b|≥$\frac{\delta}{2}$}【注：你可以画图，菱形的面积大于正方形面积】，于是 0≤P{|g(x,y)-g(a,b)|≥ε}≤P{{|x-a|≥$\frac{\delta}{2}$}∪{|y-b|≥$\frac{\delta}{2}$}}≤P{|x-a|≥$\frac{\delta}{2}$}+P{|y-b|≥$\frac{\delta}{2}$}→0，所以$\lim\limits_{n\to\infty} P\{|g(X_n, Y_n) - g(a,b)| < \varepsilon\}=1$。

切比雪夫大数定理的特殊情况：得在切比雪夫大数定理的两个条件的基础上：设随机变量序列$X_1, X_2, \cdots, X_n$独立，且有相同的数学期望和方差：E(X_i)=μ，D(X_i)=σ^2(它保证了方差存在且有共同上界)。则 1≥P{|$\bar{X}$-E($\bar{X}$)|<ε}≥1−$\frac{D(\bar{X})}{\varepsilon^2}$=1−$\frac{\sigma^2}{\varepsilon^2 n}$。所以对于任意正数$\varepsilon$，有$\lim\limits_{n\to\infty} P\{|\bar{X} - E(\bar{X})| < \varepsilon\} = \lim\limits_{n\to\infty} P\{|\bar{X} - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \varepsilon\} = \lim\limits_{n\to\infty} P\{|\bar{X} - \mu| < \varepsilon\}=1$。【相比于切比雪夫大数定理，具象化了 E($X_i$)、D($X_i$)】

伯努利大数定理(切比雪夫大数定理的更特殊情况)：设n_A为 n 次独立重复试验中，事件 A 发生的次数，p 为事件 A 在每次试验中发生的概率，则对于任意正数ε，有$\lim\limits_{n\to\infty} P\{|\frac{n_A}{n} - p| < \varepsilon\}=1$。【$\lim\limits_{n\to\infty} P\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\}=0$】——其中$X_1, X_2, \cdots, X_n$相互独立且服从参数为 p 的 0-1 分布，可知$n_A=\sum_{i=1}^{n} X_i$。则$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i=\frac{n_A}{n}$，并且之前的$\mu$变成了 p。【又把$\bar{X}$给具象化了】

我们称$\frac{n_A}{n}$为频率，伯努利大数定理表明，事件发生的频率依概率收敛于事件发生的概率。或者说频率会稳定于概率。这提供了用频率来得到概率(蒙特卡洛算法)的理论依据。

辛钦大数定理：设随机变量序列$X_1,X_2,\cdots,X_n$独立，服从同一分布，具有(当然有相同的)数学期望 $E(X_i)=\mu$。则对于任意正数ε，有$\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}=1$。

该定理相对于以前的三个，并不要求随机变量的方差存在(有共同上界)，不过需要$X_1,X_2,\cdots,X_n$服从同一分布；辛钦大数定理有三个条件，而切比雪夫大数定理只有两个，不过由于后者公式中仍存在$E(\overline{X})$，即要求数学期望存在，后者仍可看作有 3 个条件。它需要用特征函数来证明。

****

设X_i在区间[a,b]上都满足相同的均匀分布，它们间相互独立，可知它们的期望和方差均相同，满足辛钦大数定理条件，且$\overline{g(X_i)}=\frac{1}{n}\sum_{i=1}^{n}g(X_i)$，$E(\overline{g(X_i)})=\frac{1}{n}\sum_{i=1}^{n}E(g(X_i))=E(g(X_i))=\int_{-\infty}^{+\infty}g(x_i)\cdot f(x)=\int_a^b g(x)\cdot\frac{dx}{b-a}=\frac{\int_a^b g(x)\cdot dx}{b-a}$，于是根据辛钦大数定理(或者切比雪夫大数定理)，对于任意正数ε，有$\lim\limits_{n\to\infty}P\{|\overline{g(X_i)}-E(\overline{g(X_i)})|<\varepsilon\}=1$，进而$\lim\limits_{n\to\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}g(X_i)-\frac{\int_a^b g(x)\cdot dx}{b-a}|<\varepsilon\}=1$。因此我们可以利用$(b-a)\cdot\lim\limits_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}g(X_i)$来近似$\int_a^b g(x)\cdot dx$。——当 n 充分大时，结果是相当精确的。这也是为什么实际上，有些能积分的面积也不通过积分得到的原因。

5.2 中心极限定理

切比雪夫不等式告诉我们：在一定的条件下，$P\{|\overline{X}-E(\overline{X})|\geq\varepsilon\}\leq\frac{k}{\varepsilon^2 n}$。切比雪夫大数定理告诉我们，当 n 足够大时，该概率接近于 0。两者均未“精确”(但还是很“精确”了)地给出这个概率。一般来说，概率计算依赖于随机变量的概率分布，然而$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$的概率分布往往非常难求(这涉及到离散 or 连续型随机变量的卷积公式，三维的已经很难求了)，所以精确计算它又几乎是不可能的。然而马上将介绍的中心极限定理则能较好地解决这个问题。

列维-林德博格定理(独立同分布的中心极限定理)：设$X_1,X_2,\cdots,X_n$相互独立，服从同一分布，具有(相同的)数学期望和方差，则$\sum_{i=1}^{n}X_i$的标准化随机变量$(\sum_{i=1}^{n}X_i)^*=\frac{\sum_{i=1}^{n}X_i-E(\sum_{i=1}^{n}X_i)}{\sqrt{D(\sum_{i=1}^{n}X_i)}}=\frac{\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}E(X_i)}{\sqrt{\sum_{i=1}^{n}D(X_i)}}=\frac{\sum_{i=1}^{n}X_i-nE(X_i)}{\sqrt{nD(X_i)}}=\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n\sigma^2}}$，依概率$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{y^2}{2}}\cdot dy$收敛于 x。——用公式来说明即为：$\lim\limits_{n\to\infty}P\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n\sigma^2}}<x\}=\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{y^2}{2}}\cdot dy$。其中假定了$E(X_i)=\mu$、$D(X_i)=\sigma^2$。【注意：$X_i$可以在形式上为 $g(X_i)$的形式，只要是满足以上 4 个条件的随机变量即可，比如X_i^2】【这里相比于大数定理，$\overline{X}$变成了$\sum_{i=1}^{n}X_i$】

也就是说，当 n 很大时，$\sum_{i=1}^{n}X_i$的标准化变量近似地为标准正态变量，即$\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n\sigma^2}}\overset{\cdot}{\sim}N(0,1)$。其中“$\overset{\cdot}{\sim}$”表示“近似地服从”。接着根据 2.5.1 连续函数的分

布：$f_Y(y)=|(g^{-1}(y))'|\cdot f_x(g^{-1}(y))$，我们有：$X=\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n\sigma^2}}$，$Y=\sum_{i=1}^{n}X_i=\sqrt{n\sigma^2}\cdot X+n\mu$，于是$(g^{-1}(y))'=\frac{1}{\sqrt{n\sigma^2}}$、$g^{-1}(y)=\frac{y-n\mu}{\sqrt{n\sigma^2}}$，代入即有 $Y=\sum_{i=1}^{n}X_i\dot{\sim}N(n\mu,n\sigma^2)$。——同样地设 $Y=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{n}X$，则$(g^{-1}(y))'=n$，$g^{-1}(y)=ny$，则 $Y=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\dot{\sim}N(\mu,\frac{\sigma^2}{n})$。——其中，$\overline{X}\dot{\sim}N(\mu,\frac{\sigma^2}{n})$是数理统计中大样本统计推断的基础。【这里也可以用 E(Y)、D(Y)来确定 Y 的分布；Y 也服从正态分布，因为 Y 是 X 的线性组合】【它和辛钦大数定理的条件一样，也是独立同分布(+具有同数学期望)，不过添加了方差存在，比辛钦大数定理的三个条件多了一条，共 4 条；也需要用特征函数来证明】

棣莫弗-拉普拉斯定理：设随机变量$\sum_{i=1}^{n}X_i\sim B(n,p)$，其中$X_i\sim$0-1 分布，则对于任意实数 x，$\lim_{n\to\infty}P\{\frac{\sum_{i=1}^{n}X_i-nE(X_i)}{\sqrt{nD(X_i)}}<x\}=\lim_{n\to\infty}P\{\frac{\sum_{i=1}^{n}X_i-np}{\sqrt{np(1-p)}}<x\}=\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{y^2}{2}}\cdot dy$。其中$\sum_{i=1}^{n}X_i$可以用$n_A$代替。

于是类似于之前的$\sum_{i=1}^{n}X_i\dot{\sim}N(n\mu,n\sigma^2)$，用$p(1-p)$替换$\sigma^2$、用$p$替换$\mu$，便可得到$n_A\dot{\sim}N(np,np(1-p))$。

****

(1).一个系统由 n 个相互独立的部件组成，每个部件的可靠性为 0.9，至少须有 80%的部件工作才能使得整个系统正常工作。问 n 至少多大才能使系统的可靠性不低于 0.95?

设正常工作的部件数为 $X=\sum_{i=1}^{n}X_i$。其中X_i服从 0-1 分布。

$E(X)=E(\sum_{i=1}^{n}X_i)=nE(X_i)=0.9n$，$\sqrt{D(X)}=\sqrt{D(\sum_{i=1}^{n}X_i)}=\sqrt{nD(X_i)}=\sqrt{n0.9(1-0.9)}=0.3\sqrt{n}$。由于 n 较大，则$\frac{X-E(X)}{\sqrt{D(X)}}=\frac{X-0.9n}{0.3\sqrt{n}}\dot{\sim}N(0,1)$。则 $P\{X>0.8n\}=P\{\frac{X-0.9n}{0.3\sqrt{n}}>\frac{-0.1n}{0.3\sqrt{n}}\}=P\{\frac{X-0.9n}{0.3\sqrt{n}}>-\frac{\sqrt{n}}{3}\}=1-P\{\frac{X-0.9n}{0.3\sqrt{n}}\le-\frac{\sqrt{n}}{3}\}=1-\Phi(-\frac{\sqrt{n}}{3})=\Phi(\frac{\sqrt{n}}{3})\ge0.95$，查表得$\frac{\sqrt{n}}{3}\ge1.645$，得到 $n\ge25$。

但是，所谓的正确答案却是这样解的：$P\{n>X>0.8n\}=P\{\frac{0.1n}{0.3\sqrt{n}}>\frac{X-0.9n}{0.3\sqrt{n}}>\frac{-0.1n}{0.3\sqrt{n}}\}=\Phi(\frac{\sqrt{n}}{3})-\Phi(-\frac{\sqrt{n}}{3})=2\Phi(\frac{\sqrt{n}}{3})-1\ge0.95$，$\Phi(\frac{\sqrt{n}}{3})\ge0.975$，查表得$\frac{\sqrt{n}}{3}\ge1.96$，于是 $n\ge35$。

可以算算，当 $n\to\infty$时，$P\{n>X>0\}=P\{\frac{0.1n}{0.3\sqrt{n}}>\frac{X-0.9n}{0.3\sqrt{n}}>-3\sqrt{n}\}=\Phi(\frac{\sqrt{n}}{3})-\Phi(-3\sqrt{n})\to1$。这表明正常工作的部件数 X，落在[0,+∞)间的概率为 1。

(2).某电器元件寿命服从均值为 100h 的指数分布，随机从中取 16 只，求 16 只元件的寿命总和>1920h 的概率?

设寿命总和为 $X=\sum_{i=1}^{16}X_i$。其中X_i服从参数λ为 0.01 的指数分布。

$E(X)=nE(X_i)=n\cdot\frac{1}{\lambda}=16\cdot100=1600$，$\sqrt{D(X)}=\sqrt{nD(X_i)}=\sqrt{n\cdot\frac{1}{\lambda^2}}=4\cdot100=400$，由于 n=16 较大，于是$\frac{X-E(X)}{\sqrt{D(X)}}=\frac{X-1600}{400}\dot\sim N(0,1)$，则 $P\{X>1920\}=P\{\frac{X-1600}{400}>0.8\}=1-\Phi(0.8)=1-0.7881=0.2119$。

(3).200 台车床，开工率 0.6，每台车床独立工作，开工需要 1kw，多少瓦能以 99.9%的概率保证车间不会因供电不足影响生产?

设正常工作的台数为 $X=\sum_{i=1}^{200}X_i$。其中X_i服从 0-1 分布。设 Z 千瓦可以保证。

$E(X)=nE(X_i)=120$，$\sqrt{D(X)}=\sqrt{nD(X_i)}=\sqrt{120\cdot0.4}=4\sqrt{3}$，由于 n=200 较大，于是$\frac{X-E(X)}{\sqrt{D(X)}}=\frac{X-120}{4\sqrt{3}}\dot\sim N(0,1)$，$P\{0<X<Z/1\}=P\{\frac{-120}{4\sqrt{3}}<\frac{X-120}{4\sqrt{3}}<\frac{Z-120}{4\sqrt{3}}\}=\Phi\left(\frac{Z-120}{4\sqrt{3}}\right)-\Phi\left(\frac{-120}{4\sqrt{3}}\right)>0.999$，其中红色部分可以去掉，因为题目只涉及不影响生产，而没有其他条件限制(虽然从实际上确实也有这样的条件限制)。【其中 Z/1=Zkw/1kw=所供的功率所能保证正常生产的台数】

(4).0~9 中取多少次，才能使得“0”出现的频率在 0.09~0.11 之间的概率为 0.95?

设 0 出现的次数为 $X=\sum_{i=1}^{n}X_i$。其中X_i服从 0-1 分布(出现 0 记为 1)。

$E(X)=nE(X_i)=0.1n$，$\sqrt{D(X)}=\sqrt{nD(X_i)}=\sqrt{n\cdot0.1\cdot0.9}=0.3\sqrt{n}$，由于 n 较大，于是$\frac{X-E(X)}{\sqrt{D(X)}}=\frac{X-0.1n}{0.3\sqrt{n}}\dot\sim N(0,1)$，$P\{0.09<\frac{X}{n}<0.11\}=P\{0.09n<X<0.11n\}=P\{\frac{0.09n-0.1n}{0.3\sqrt{n}}<\frac{X-0.1n}{0.3\sqrt{n}}<\frac{0.11n-0.1n}{0.3\sqrt{n}}\}=P\{\frac{-0.01n}{0.3\sqrt{n}}<\frac{X-0.1n}{0.3\sqrt{n}}<\frac{0.01n}{0.3\sqrt{n}}\}=\Phi\left(\frac{\sqrt{n}}{30}\right)-\Phi\left(-\frac{\sqrt{n}}{30}\right)=2\Phi\left(\frac{\sqrt{n}}{30}\right)-1>0.95$，即有$\Phi\left(\frac{\sqrt{n}}{30}\right)>0.975$，$\frac{\sqrt{n}}{30}>1.96$，于是 $n\geq3458$。

数理统计

第六章 样本与抽样分布

前五章属于概率论范畴，初步了解了事件的概率和随机变量。许多实际问题中的随机现象都可以用随机变量来描述，而要很好地了解一个随机变量，就需要知道它的概率分布，至少也要知道它的数字特征。

在概率论中，一切的分析和运算都是基于分布是已知这个假设进行的。但是在实际问题中，我们常常对所要研究的随机变量知之甚少。这时便需要试验和观测，获

得反映随机变量信息的数据，并以概率论为理论基础，对数据进行整理、分析和推断，进而对研究对象的性质和统计规律做出合理科学的估计和推断。这就是数理统计的基本的和主要的任务。

数理统计的应用领域很广泛，是众多学科从事科学研究和生产必不可少的工具。

6.1 总体与样本

6.1.1 总体和个体

数理统计中，通常把研究对象的全体称为总体，将总体中的每个元素称为个体。我们往往关心的是个体的某项指标，而不是个体的一切属性。对于选定的某个数量指标 X 而言，每个个体的指标取值是确定的，但是在抽取之前 X 的取值又是无法预测的。

因此数量指标 X 是个随机变量，它的分布反映了总体在这个数量指标上的分布规律。从只关心总体的数量指标的分布规律来看，今后我们就把数量指标与总体等同看待。即将研究对象的某项数量指标看作总体，把数量指标的每个取值作为个体。

若我们需要对某一对象的多个数量指标进行研究，这时可采用多维随机变量及其分布来描述总体，此时总体称为多维总体。总体中所包含的个体的数量称为总体容量。根据总体容量的有限 or 无限，分为有限总体和无限总体。

6.1.2 样本和简单随机样本

人们通常通过抽样的方法了解总体分布。总体中抽取出的一部分个体，称为总体的一个样本，样本中的个体称为样品，样本中做包含的个体的数目称为样本容量。【与第一章的联系：每一个样品本身可看作一个样本空间，包含了所有试验可能的结果，而每个结果为一个样本点。这里，“总体”相当于“样本空间”的地位，“样本”的地位相当于之前“事件”的地位；“样品”相当于“样本点”的地位；总体>样本>样品=样本空间>事件>样本点，其中，前三个属于随机变量范畴，后三个属于随机变量的取值范畴】

由于要通过样本来了解总体，所以样本得具有代表性，这就对样本的抽取方式提出了要求。最常用的抽取方法为：在相同条件下对总体 X 进行 n 次重复且独立的随

机观测，并把 n 次观测结果记为$X_1,X_2,\cdots,X_n$。采取这种有放回抽取得到的样本$X_1,X_2,\cdots,X_n$是互相独立的随机变量，且这些样品X_i与总体 X 有相同的分布，因此该样本具有代表性，这样的样本称为简单随机样本。样本的一次观测值，即其中所有样品的各自的一个取值的排列，记为$x_1,x_2,\cdots,x_n$，称为样本的一组样本值。【以前的$x_1,x_2,\cdots,x_n$表示的是某个X_i的所有样本点，现在是取尽每个X_i的一个样本点出来组成的序列】其中的 n 就是样本容量。——以后未特别说明，今后的讨论中所提到的样本均指简单随机样本。

采用有放回抽取才能得到简单随机样本。当总体容量比样本容量大得多时，可将无放回抽取方式所得到样本当作简单随机样本。对无限总体来说，无放回抽取方式所得到样本就是简单随机样本，因抽走少量样本后不影响总体的构成或影响很小。

6.1.3 样本的联合分布

(不管$X_1,X_2,\cdots,X_n$或者说 X 连续还是离散)若总体 X 的分布函数为$F_X(x)$，那么样本$X_1,X_2,\cdots,X_n$的联合分布函数为 $F(x_1,x_2,\cdots,x_n)=\prod_{i=1}^n F_X(x_i)$【这是因为$X_1,X_2,\cdots,X_n$互相独立】。

当 X 为连续型随机变量且概率密度函数为$f_X(x)$时，样本$X_1,X_2,\cdots,X_n$的联合密度函数为 $f(x_1,x_2,\cdots,x_n)=\prod_{i=1}^n f_X(x_i)$。

当 X 为离散型随机变量且概率分布为 $P\{X=x\}$时，样本$X_1,X_2,\cdots,X_n$的联合概率分布为 $f(x_1,x_2,\cdots,x_n)=P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^n P\{X_i=x_i\}$。

6.1.4 经验分布函数

设总体 X 的一组样本值$x_1,x_2,\cdots,x_n$。将其由小到大排列，重复的值相邻地放在一起，比如其中有n_1个最小值$x_i,x_j,\cdots,x_k$，将它们全部映射为相同个数的$x_{(1)},x_{(1)},\cdots,x_{(1)}$新序列，使得$x_{(1)}<x_{(2)}<\cdots<x_{(m)}$，且$n_1+n_2+\cdots+n_m=n$。其中的$x_{(i)}$也可用$x_i^*$表示。

定义$F_n(x)=0,x<x_{(1)}$； $=\frac{1}{n}\sum_{i=1}^k n_i,x_{(k)}\le x<x_{(k+1)}(k=1,2,\cdots,m\text{-}1)$； $1,x_{(m)}\le x$ 为总体 X 关于样本值$x_1,x_2,\cdots,x_n$的经验分布函数 or 样本分布函数$F_n(x)$。

格利文科定理：对于任意实数 x，经验分布函数$F_n(x)$以概率 1 一致收敛于总体分布函数$F(x)$：$\lim_{n\to\infty}\sup_{-\infty<x<+\infty} P\{|F_n(x)-F(x)|=0\}=1$。【证明用到特征函数】

*统计，是从手中已有资料(样本值)，去推断总体的分布情况(即 F(x)的性质)。

6.2 样本的数字特征

6.2.1 样本的基本数字特征

不含任何未知参数的，样本的函数 $g(X_1,X_2,\cdots,X_n)$，叫统计量。是完全能由样本决定的量。统计量依赖于样本，而样本又是随机变量组，所以统计量也是随机变量。$g(x_1,x_2,\cdots,x_n)$称为统计量的观测值。

几个常见的统计量：

(1).样本均值：$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$。其中$X_i$为样品。

(2).样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$。

为什么是$\frac{1}{n-1}$而不是$\frac{1}{n}$呢？这是因为其期望：$E(S^2)=\frac{1}{n-1}E[\sum_{i=1}^{n}(X_i-\overline{X})^2]=\frac{1}{n-1}\sum_{i=1}^{n}E[(X_i-\overline{X})^2]=\frac{1}{n-1}\sum_{i=1}^{n}E[X_i^2-2X_i\overline{X}+\overline{X}^2]=\frac{1}{n-1}\sum_{i=1}^{n}\{E[X_i^2]-2E[X_i\overline{X}]+E[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-2\sum_{i=1}^{n}(E[X_i]E[\overline{X}])+nE[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-2E[\overline{X}]\sum_{i=1}^{n}E[X_i]+nE[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-2E[\overline{X}]E[\sum_{i=1}^{n}X_i]+nE[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-2E[\overline{X}]E[n\overline{X}]+nE[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-2nE[\overline{X}]E[\overline{X}]+nE[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}(D[X_i]+E^2[X_i])-2nE^2[\overline{X}]+n(D[\overline{X}]+E^2[\overline{X}])\}=\frac{1}{n-1}\{\sum_{i=1}^{n}(D[X_i]+E^2[X_i])+nD[\overline{X}]-nE^2[\overline{X}])\}$

Oh 不，我们不能使用$E[X_i\overline{X}]=E[X_i]E[\overline{X}]$，因为$X_i$和$\overline{X}$不是独立的。不过我们可以类似之后$\sum_{i=1}^{n}E[X_i]=E[\sum_{i=1}^{n}X_i]$的地，学习着在这一步就开始这样做手脚：$E(S^2)=\frac{1}{n-1}E[\sum_{i=1}^{n}(X_i-\overline{X})^2]=\frac{1}{n-1}\sum_{i=1}^{n}E[(X_i-\overline{X})^2]=\frac{1}{n-1}\sum_{i=1}^{n}E[X_i^2-2X_i\overline{X}+\overline{X}^2]=\frac{1}{n-1}\sum_{i=1}^{n}\{E[X_i^2]-2E[X_i\overline{X}]+E[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-2E[\sum_{i=1}^{n}(X_i\overline{X})]+nE[\overline{X}^2]\}$，其中$E[\sum_{i=1}^{n}(X_i\overline{X})]=E[\overline{X}\sum_{i=1}^{n}(X_i)]=E[\overline{X}n\overline{X}]=nE[\overline{X}^2]$，因此$E(S^2)=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-2nE[\overline{X}^2]+nE[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}E[X_i^2]-nE[\overline{X}^2]\}=\frac{1}{n-1}\{\sum_{i=1}^{n}(D[X_i]+E^2[X_i])-n(D[\overline{X}]+E^2[\overline{X}])\}$

由于其中$D[X_i]=D[X]=\sigma^2$，$D[\overline{X}]=D[\frac{1}{n}\sum_{i=1}^{n}X_i]=\frac{1}{n^2}D[\sum_{i=1}^{n}X_i]=\frac{1}{n^2}\sum_{i=1}^{n}D[X_i]=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$，且$E^2[X_i]=E^2[X]=\mu^2$，$E^2[\overline{X}]=E^2[\frac{1}{n}\sum_{i=1}^{n}X_i]=\mu^2$，此时$E(S^2)=\frac{1}{n-1}\{\sum_{i=1}^{n}(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\}=\frac{1}{n-1}\{(n\sigma^2+n\mu^2)-(\sigma^2+n\mu^2)\}=\sigma^2$，即仅仅为了保证$E(S^2)=\sigma^2$而创造的系数。【注：可以证明，将恒有$E(S^2)=D[X_i]=D[X]$；利用其中所体现的以下结论：$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$】

 chzh_xie@foxmail.com. (contact)

(3).样本 k 阶原点矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$。——和之前的 X 的 k 阶原点矩 $E(X^k)$不同，虽然$\frac{1}{n}\sum_{i=1}^{n}(\cdot)$相当于 $E(\cdot)$，但这里的结果$\frac{1}{n}\sum_{i=1}^{n}X_i^k$是个变量，而 $E(X^k)$的结果是个数值。

(4).样本 k 阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$。——也和之前的 X 的 k 阶中心矩 $E([X-E(X)]^k)$不同，$E(X)$是个数，而$\overline{X}$是个变量，因而之前在求 $E(S^2)$时，不能沿用之前的思路把$\overline{X}$当作个数提取到 E()外。——B_2和S^2很像，且$B_2=\frac{n-1}{n}S^2$。

6.3 三个常用的抽样分布

卡方分布、t 分布、F 分布，都是由正态分布导出的分布，是试验统计中常用的分布。在数理统计中占有十分重要的地位。

(一).三大分布

1.χ^2分布(卡方分布)：$\chi^2(n)$

若$X_1,X_2,\cdots,X_n$都相互独立且都服从 N(0,1)，则 $X=X_1^2+X_2^2+\cdots+X_n^2$所服从的分布称为自由度为 n 的$\chi^2$分布，记作 $X\sim\chi^2(n)$。

它的概率密度函数为 $f(x)=\frac{1}{2^{\frac{n}{2}}\cdot\Gamma(\frac{n}{2})}\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}$,x>0； =0,else。其中 $\Gamma(\alpha)=\int_0^{+\infty}t^{\alpha-1}\cdot e^{-t}dt$。Γ函数的基本性质：①.$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$ ②.对于正整数 n，$\Gamma(n)=(n-1)!$ ③.$\Gamma(0.5)=\sqrt{\pi}$，前两点性质我们已经在常用的二维连续随机变量的函数分布中给出过证明，下面对第三点给出证明：$\Gamma(0.5)=\int_0^{+\infty}t^{-0.5}\cdot e^{-t}dt$，令 $t=\frac{S^2}{2}$，则 $=\sqrt{2}\int_0^{+\infty}S^{-1}\cdot e^{-\frac{S^2}{2}}d(\frac{S^2}{2})=\sqrt{2}\int_0^{+\infty}e^{-\frac{S^2}{2}}dS=2\sqrt{\pi}\int_0^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{S^2}{2}}dS=2\sqrt{\pi}\cdot 0.5=\sqrt{\pi}$。

我们来利用数学归纳法证明一下它的概率密度函数是这个。首先我们看看当 $X=X_1^2$时，是否有：$X\sim\Gamma(\frac{1}{2},2)$。其中 $X\sim\Gamma(\alpha,\beta)$满足α,β>0，$f(x)=\frac{1}{\beta^\alpha\Gamma(\alpha)}\cdot x^{\alpha-1}\cdot e^{-\frac{x}{\beta}}$，x>0；0，else。——我们在 2.5.1 连续函数的分布给出了 $X=X_1^2$时的密度函数：$f_X(x)=\frac{1}{\sqrt{2\pi}}\cdot x^{-0.5}\cdot e^{-\frac{x}{2}}$，x>0；0，x≤0。它符合$\Gamma(\frac{1}{2},2)=\frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})}\cdot x^{\frac{1}{2}-1}\cdot e^{-\frac{x}{2}}=\frac{1}{\sqrt{2\pi}}\cdot x^{-0.5}\cdot e^{-\frac{x}{2}}$，x>0；0，else。

同样，我们看看当 $X=X_1^2+X_2^2$时，X 会服从参数α,β为何值的Γ分布，我们在 3.6 二维随机变量的函数的分布给出了 $X=X_1^2+X_2^2$的密度函数：$f_X(x)=\frac{1}{2}e^{-\frac{x}{2}}$，x>0。发现它

符合$\Gamma(\frac{2}{2},2)=\frac{1}{2^1\Gamma(1)}\cdot x^{1-1}\cdot e^{-\frac{x}{2}}=\frac{1}{2}e^{-\frac{x}{2}}$，x>0；0，else。因此我们推测 X=$X_1^2+X_2^2+\cdots+X_n^2$服从$\Gamma(\frac{n}{2},2)$。

又因我们在之前 3.6 中证明了Γ分布的性质：若$X_1\sim\Gamma(\alpha_1,\beta)$、$X_2\sim\Gamma(\alpha_2,\beta)$，且它们互相独立，则$X_1+X_2\sim\Gamma(\alpha_1+\alpha_2,\beta)$。则我们通过证明 X=$X_1^2$时，X~$\Gamma(\frac{1}{2},2)$以及假设 X=$X_1^2+X_2^2+\cdots+X_n^2\sim\Gamma(\frac{n}{2},2)$，会利用此性质得到 X=$X_1^2+X_2^2+\cdots+X_n^2+X_{n+1}^2\sim\Gamma(\frac{n}{2}+\frac{1}{2},2)$即$\Gamma(\frac{n+1}{2},2)$。与假设无矛盾，因此 X~$\Gamma(\frac{n}{2},2)$。而$\Gamma(\frac{n}{2},2)$就是其概率密度函数$\frac{1}{2^{\frac{n}{2}}\cdot\Gamma(\frac{n}{2})}\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}$。

卡方分布的性质：

(1).若 X~$\chi^2(n_1)$、Y~$\chi^2(n_2)$，且它们相互独立，则$X+Y\sim\chi^2(n_1+n_2)$。

这是因为：若 X~$\chi^2(n_1)$、Y~$\chi^2(n_2)$，即$X\sim\Gamma(\frac{n_1}{2},2)$、$Y\sim\Gamma(\frac{n_2}{2},2)$，且它们互相独立，则$X+Y\sim\Gamma(\frac{n_1}{2}+\frac{n_2}{2},2)$，即$X+Y\sim\Gamma(\frac{n_1+n_2}{2},2)$，即$X+Y\sim\chi^2(n_1+n_2)$。

(2).若 X~$\chi^2(n)$，则 E(X)=n，D(X)=2n。

证明：E(X)=E($\sum_{i=1}^n X_i^2$)=$\sum_{i=1}^n E(X_i^2)=\sum_{i=1}^n(D[X_i]+E^2[X_i])=\sum_{i=1}^n(1+0)$=n。之前我们在 4.1.2 连续型随机变量的数学期望也用过定义的方法求它。

D(X)=D($\sum_{i=1}^n X_i^2$)=$\sum_{i=1}^n D(X_i^2)=n\{E(X_i^4)-E^2[X_i^2]\}=n\{E(X_i^4)-1\}$，而$E(X_i^4)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot x^4\cdot e^{-\frac{x^2}{2}}\cdot dx=-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot x^3\cdot e^{-\frac{x^2}{2}}\cdot d(-\frac{x^2}{2})=-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot x^3\cdot d(e^{-\frac{x^2}{2}})=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}\cdot d(x^3)=3\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\cdot x^2\cdot e^{-\frac{x^2}{2}}\cdot dx=3E(X_i^2)=3$。因此 D(X)=$n\{E(X_i^4)-1\}=2n$。

****

串联等电阻总功率就服从卡方分布。

2.t分布：t(n)

若X与 Y 都相互独立且 X~N(0,1)、Y~$\chi^2(n)$，则 T=$\frac{X}{\sqrt{Y/n}}$所服从的分布称为自由度为 n 的t分布，记作 T~t(n)。

它的概率密度函数为 f(x)=$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\cdot\Gamma(\frac{n}{2})}\cdot(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$，-∞<x<+∞。【其中的 x 相当于 t，X 相当于 T】其密度函数关于 y 轴对称，且 n 很大时与标准正态分布的密度函数曲线几乎重叠。

我们来利用 3.6 二维随机变量的函数的分布中的雅可比变换证明一下它的概率密度函数是这个：由于 X、Y 相互独立，则 $(X,Y)\sim f_{X,Y}(x,y)=f_X(x)\cdot f_Y(y)=C\cdot e^{-\frac{x^2}{2}}\cdot y^{\frac{n}{2}-1}\cdot e^{-\frac{y}{2}}$，令 $x=r\sin\theta$，$y=r^2\cos^2\theta$，其中 $r\in(0,+\infty)$，$\theta\in(-\frac{\pi}{2},\frac{\pi}{2})$，$C=\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{2^{\frac{n}{2}}\cdot\Gamma(\frac{n}{2})}$，

则 $dxdy=\begin{vmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial\theta}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial\theta}\end{vmatrix}drd\theta=\begin{vmatrix}\sin\theta & r\cos\theta\\ 2r\cos^2\theta & -r^2\sin2\theta\end{vmatrix}drd\theta=(-2r^2\sin^2\theta\cos\theta-2r^2\cos^2\theta\cos\theta)drd\theta=-2r^2\cos\theta drd\theta$。

而 $(R,\theta)\sim f_{X,Y}(x(r,\theta),y(r,\theta))=f_{R,\theta}(r,\theta)=C\cdot e^{-\frac{r^2}{2}}\cdot(r^2\cos^2\theta)^{\frac{n}{2}-1}=C\cdot e^{-\frac{r^2}{2}}\cdot r^{n-2}\cos^{n-2}\theta$。于是 $F_T(t)=P\{T\le t\}=P\{g(X,Y)\le t\}=\iint_{g(x,y)\le t}f_{X,Y}(x,y)\cdot dxdy$ $=\iint_{g(x(r,\theta),y(r,\theta))\le t}f_{X,Y}(x(r,\theta),y(r,\theta))\cdot dxdy=\iint_{g(x(r,\theta),y(r,\theta))\le t}C\cdot e^{-\frac{r^2}{2}}\cdot r^{n-2}\cos^{n-2}\theta\cdot -2r^2\cos\theta drd\theta=\iint_{g(x(r,\theta),y(r,\theta))\le t}-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cos^{n-1}\theta\cdot drd\theta$，其中 $f_{R,\theta}(r,\theta)=-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cos^{n-1}\theta$。又因 $t=\frac{x}{\sqrt{y/n}}=\sqrt{n}\tan\theta$，因此 $\sec\theta=\sqrt{(1+\tan^2\theta)}=(1+\frac{t^2}{n})^{\frac{1}{2}}$，而 $dt=\sqrt{n}\sec^2\theta=\sqrt{n}(1+\frac{t^2}{n})d\theta$，所以 $d\theta=[\sqrt{n}(1+\frac{t^2}{n})]^{-1}dt$，将 $d\theta$ 换成 dt，且将 $\cos\theta$ 换成 $(1+\frac{t^2}{n})^{-\frac{1}{2}}$ 地代入 $F_T(t)=\iint_{g(x(r,\theta),y(r,\theta))\le t}-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cos^{n-1}\theta\cdot drd\theta=$ $\iint_{g(x(r,\theta(t)),y(r,\theta(t)))\le t}-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cdot dr\cdot(1+\frac{t^2}{n})^{-\frac{n-1}{2}}\cdot[\sqrt{n}(1+\frac{t^2}{n})]^{-1}dt$ $=\iint_{g(x(r,\theta(t)),y(r,\theta(t)))\le t}-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cdot dr\cdot\frac{1}{\sqrt{n}}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}\cdot dt$。其中 $f_{R,T}(r,t)=-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cdot\frac{1}{\sqrt{n}}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}$。

从同样描述 t 的分布的，由密度函数 $f_{X,Y}(x,y)$ 转变为的密度函数 $f_{R,\theta}(r,\theta)=f_{X,Y}(x(r,\theta),y(r,\theta))\cdot-2r^2\cos\theta=-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cos^{n-1}\theta$ 中，我们可看到 R、θ 相互独立；因而再从密度函数 $f_{R,\theta}(r,\theta)$ 转变为的密度函数 $f_{R,T}(r,t)=f_{R,\theta}(r,\theta(t))\cdot\frac{d\theta}{dt}=-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cdot\frac{1}{\sqrt{n}}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}$ 中，也有 R、T 相互独立。

所以利用此性质，由于 $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_0^{+\infty}f_{R,\theta}(r,\theta)\cdot drd\theta=1$，所以 $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^{n-1}\theta\,d\theta\int_0^{+\infty}(-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n)\cdot dr=1$，得到 $\int_0^{+\infty}(-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n)\cdot dr=\frac{1}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^{n-1}\theta d\theta}$ $=\frac{1}{\frac{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})}}=\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}$。同样的道理，$\int_0^{+\infty}\int_{-\infty}^{+\infty}f_{R,T}(r,t)\cdot drdt=1$，于是 $\int_0^{+\infty}-2C\cdot e^{-\frac{r^2}{2}}\cdot r^n\cdot dr\int_0^{+\infty}\frac{1}{\sqrt{n}}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}dt=1$，因此 $\frac{1}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^{n-1}\theta d\theta}\int_0^{+\infty}\frac{1}{\sqrt{n}}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}dt=1$（这一步可以不要），即 $\int_0^{+\infty}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}\frac{1}{\sqrt{n}}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}dt=1$，即 $\int_0^{+\infty}\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\cdot\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}dt=1$。

t 分布的性质：

(1).若 T~t(n)，则 E(T)=0，D(T)=$\frac{n}{n-2}$。

证明：E(T)=$E(\frac{X}{\sqrt{Y/n}})=E(X\cdot\frac{\sqrt{n}}{\sqrt{Y}})=E(X)\cdot E(\frac{\sqrt{n}}{\sqrt{Y}})=0\cdot E(\frac{\sqrt{n}}{\sqrt{Y}})$=0。【注：X 与 Y 相互独立则X与$\frac{\sqrt{n}}{\sqrt{Y}}$也相互独立】——同样，你也可以利用t(n)分布的概率密度函数$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\cdot\Gamma(\frac{n}{2})}\cdot(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$为偶函数的性质，通过定义法来求 E(T)=对奇函数 xf(x)的积分=0。

D(T)=$E(\frac{X^2}{Y/n})-E^2(T)=E(X^2)\cdot E(\frac{n}{Y})=1\cdot nE(\frac{1}{Y})$。而$E(\frac{1}{Y})=\int_0^{+\infty}\frac{1}{x}\cdot\frac{1}{2^{\frac{n}{2}}\cdot\Gamma(\frac{n}{2})}\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}\cdot dx=\frac{1}{n-2}\int_0^{+\infty}\frac{1}{2^{\frac{n}{2}-1}\cdot\Gamma(\frac{n}{2}-1)}\cdot x^{\frac{n}{2}-2}\cdot e^{-\frac{x}{2}}\cdot dx=\frac{1}{n-2}$。(同样利用$\Gamma(\frac{n}{2}+1)=\frac{n}{2}\cdot\Gamma(\frac{n}{2})$所对应的$\Gamma(\frac{n}{2})=(\frac{n}{2}-1)\cdot\Gamma(\frac{n}{2}-1)$)。因此有 D(T)=$nE(\frac{1}{Y})=\frac{n}{n-2}$。

【常见的一些爱考的服从 t 分布的统计量：1.设总体 X~N(0,σ^2)，其中得有σ^2=1，则$\frac{X_1-X_2}{\sqrt{X_3^2+X_4^2}}$~t(2)；而$\frac{X_1+X_2}{\sqrt{X_1^2+X_2^2}}$由于不独立而并不服从 t 分布。】

3.F 分布：F(n_1,n_2)

若X与 Y 都相互独立且 X~$\chi^2(n_1)$、Y~$\chi^2(n_2)$，则 F=$\frac{X/n_1}{Y/n_2}$所服从的分布称为第一自由度为n_1、第二自由度为n_2的 F 分布，记作 F~F(n_1,n_2)。

它的概率密度函数为 f(x)=$\frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_2}{2})\cdot\Gamma(\frac{n_1}{2})}\cdot\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}x^{\frac{n_1}{2}-1}\cdot(1+\frac{n_1}{n_2}x)^{-\frac{n_1+n_2}{2}}$，x>0。

F 分布的性质：

(1).若 F~F(n_1,n_2)，则$\frac{1}{F}$~F(n_2,n_1)。

这是因为$\frac{1}{F}=\frac{Y/n_2}{X/n_1}$，且 Y~$\chi^2(n_2)$、X~$\chi^2(n_1)$、X与 Y 相互独立。

(二).上α分位数

1.若 X~$\chi^2(n)$，任给一α∈(0,1)，使得 P{X>$\chi^2(n)$}=α成立的$\chi^2(n)$记为$\chi^2_\alpha(n)$，称为自由度为 n 的χ^2分布的上α分位数。【这里的$\chi^2(n)$相当于小 x；另外，这非常类似极限的定义中的“反函数/反向映射”的关系：正如已知ε，去求 N(ε)一样；这里也是，已知某点横坐标到右无穷所 cover 的概率/面积α，去求(查表)对应的横坐标$\chi^2_\alpha(n)$】

由于$X_1,X_2,\cdots,X_n$均独立同分布，则$X_1^2,X_2^2,\cdots,X_n^2$也都独立同分布，且具有相同的期望和方差，根据中心极限定理，有$\lim\limits_{n\to\infty}P\{\frac{\sum_{i=1}^n X_i^2-nE(X_i^2)}{\sqrt{nD(X_i^2)}}<x\}=\lim\limits_{n\to\infty}P\{\frac{\sum_{i=1}^n X_i^2-E(\sum_{i=1}^n X_i^2)}{\sqrt{D(\sum_{i=1}^n X_i^2)}}<x\}=\Phi(x)$，即$\frac{X-n}{\sqrt{2n}}\dot\sim N(0,1)$。于是$\frac{\chi^2_\alpha(n)-n}{\sqrt{2n}}\approx z_\alpha$，即$\chi^2_\alpha(n)\approx n+\sqrt{2n}\cdot z_\alpha$。——其中$z_\alpha$为标准正

态分布的上α分位数。【关于z_α与$\Phi(x)$的联系：在标准正态分布函数值表中，是通过已知$\Phi(x)$的值，去查对应的未知的x的值；而在诸如χ^2分布表、t 分布表中，是通过已知α的值，去查对应的未知的z_α的值——二者虽然一个($\Phi(x)$)是通过因变量$\Phi(x)$去找自变量x(反向)，另一个(z_α)是通过自变量α去找因变量z_α(正向)，但都是通过面积/概率去找对应的横坐标；只不过对于面积，一个($\Phi(x)$)是从-∞积分到x，另一个(z_α中的α)是从z_α积分到+∞。——因此很容易有：$\Phi(z_\alpha)=1-\alpha$、$\Phi(z_{1-\alpha})=\alpha$，如果我们能创造一个像$z_\alpha$的函数，命名为$\Phi_\alpha$或者$x_\Phi$，即有：$\Phi_\alpha=z_{1-\alpha}$，或者$x_\alpha=z_{1-\alpha}$】

由于χ^2分布的密度函数中出现了Γ函数：$\Gamma(\frac{n}{2})$，直接求$\chi^2_\alpha(n)$很困难，因此对于不同的 n 和α，上α分位数的值已制成了χ^2分布表。当 n>45 时，$\chi^2(n)$分布近似于正态分布【但不一定是标准正态分布，$X\dot{\sim}N(n,2n)$】，所以可以利用标准正态分布的上α分位数z_α近似得到$\chi^2_\alpha(n)$【通过$\chi^2_\alpha(n)\approx n+\sqrt{2n}\cdot z_\alpha$，即通过$\chi^2_\alpha(n)\approx E(\sum_{i=1}^n X_i^2)+\sqrt{D(\sum_{i=1}^n X_i^2)}\cdot z_\alpha$，即通过$E(X)+\sqrt{D(X)}\cdot z_\alpha$，其中 $X\sim\chi^2(n)$】，此时表中 n 止于 45。

2.若 $T\sim t(n)$，则使得 $P\{X>t_\alpha(n)\}=\alpha$成立的$t_\alpha(n)$称为自由度为 n 的t分布的上α分位数。

t(n)分布的上α分位数的值也已制成了 t 分布表。仍只有 n<45 时的$t_\alpha(n)$值。当 n>45 时，t(n)分布近似于标准正态分布，此时$t_\alpha(n)\approx z_\alpha$。【这一点可以通过 t 分布的密度函数得知，或者从 E(T)=0，$D(T)=\frac{n}{n-2}$看出端倪；或者通过$T=\frac{X}{\sqrt{Y/n}}\dot{\sim}\frac{N(0,1)}{\sqrt{N(n,2n)/n}}=\frac{N(0,1)}{\sqrt{N(1,2/n)}}$? 】

额外地，由t(n)分布的对称性可知：$t_{1-\alpha}(n)=-t_\alpha(n)$。【对于标准正态分布，也有$z_{1-\alpha}=-z_\alpha$】

3.$F\sim F(n_1,n_2)$，则使得 $P\{F>F_\alpha(n_1,n_2)\}=\alpha$成立的$F_\alpha(n_1,n_2)$称为第一自由度为$n_1$、第二自由度为$n_2$的 F 分布的上α分位数。

F 分布的上α分位数的值也已制成了 F 分布表。但表中只给出了一些常用的α值所对应的上α分位数，而上 1-α分位数可由下式得到：$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$。

其证明过程：由于$\frac{1}{F}\sim F(n_2,n_1)$，$\alpha=P\{F>F_\alpha(n_1,n_2)\}=P\{\frac{1}{F}>F_\alpha(n_2,n_1)\}=P\{F<\frac{1}{F_\alpha(n_2,n_1)}\}=1-P\{F\geq\frac{1}{F_\alpha(n_2,n_1)}\}=1-P\{F>\frac{1}{F_\alpha(n_2,n_1)}\}$。于是 $P\{F>\frac{1}{F_\alpha(n_2,n_1)}\}=1-\alpha$。又因$1-\alpha=P\{F>F_{1-\alpha}(n_1,n_2)\}$。因此 $P\{F>\frac{1}{F_\alpha(n_2,n_1)}\}=P\{F>F_{1-\alpha}(n_1,n_2)\}$，即有$\frac{1}{F_\alpha(n_2,n_1)}=F_{1-\alpha}(n_1,n_2)$。【有个对此的记忆方式非常不错：“三倒”——倒数、位置交换、相加等于 1】

【常见的一些爱考的服从 F 分布的统计量：1.设总体 X~N(0,1)，则 $\frac{(X_1-X_2)^2}{(X_1+X_2)^2}$~F(1,1)(其中 cov($X_1-X_2$,$X_1+X_2$)=0 能说明$X_1-X_2$与$X_1+X_2$相互独立? ——若它们是正态变量，则可以说明)、$\frac{(X_1-X_2)^2}{(X_3+X_4)^2}$~F(1,1)。】

6.4 常用统计量及其分布

样本是总体的代表，包含总体的许多信息，这些信息比较分散，不能直接用于对总体进行统计推断。得进行整理、加工，将有用信息集中起来。以样本作为自变量的样本函数就是样本信息集中的一种表现，我们可利用它对总体进行统计推断。

设$X_1,X_2,\cdots,X_n$是总体的一个样本，样本函数 g($X_1,X_2,\cdots,X_n$)，叫统计量。统计量是随机变量。6.2 中介绍的样本的数字特征都是统计量。

以下两条定理均设$X_1,X_2,\cdots,X_n$是来自总体 N(μ,σ^2)的一个样本(样品之间都是互相独立的)。

1.定理 1：关于样本均值$\overline{X}$

(1).$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$。

(2).$\overline{X}^*=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$。 【这一点和定理 2 的(3).相联系，7.2 会用上】

(3).$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$。 【这一点和定理 2 的(1).相联系，或许这些东西就是为了 7.2 而开发出来的】

据此我们可以得出一些有用的东西：$E(\sum_{i=1}^{n}(X_i-\mu)^2)=\sigma^2E(\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2)=\sigma^2\cdot n$，则【或许更像离散总体 X 的 k 阶中心矩】，E(像B_2)=$E(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2)=\sigma^2$；同样，$D(\sum_{i=1}^{n}(X_i-\mu)^2)=\sigma^4D(\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2)=\sigma^4\cdot 2n$；因而 $D(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2)=\frac{1}{n^2}\sigma^4\cdot 2n=\frac{2\cdot\sigma^4}{n}$。

2.定理 2：关于样本方差S^2

(1).$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$。

虽然看上去$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$，是 n 个随机变量的平方和，那么$\frac{(n-1)S^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}$该服从$\chi^2(n)$才对。——但是这 n 个变量中只有 n-1 个是线性无关的，因此独立的 n-1 个对应的是$\chi^2(n-1)$。我们来证明它：

由于$X_1,X_2,\cdots,X_n$相互独立，则它们的联合密度函数$=(\frac{1}{\sqrt{2\pi}\sigma})^n\cdot e^{-\frac{\sum_i(x_i-\mu)^2}{2\sigma^2}}=(\frac{1}{\sqrt{2\pi}\sigma})^n\cdot e^{-\frac{\sum_i x_i^2-2\mu\sum_i x_i+n\mu^2}{2\sigma^2}}$。令(存在这样的正交阵)某一正交矩阵 $\mathbf{A}(\mathbf{A}\mathbf{A}^T=\mathbf{A}^T\mathbf{A}=\mathbf{E})$，满足 $\mathbf{Y}=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=\mathbf{A}\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}=\mathbf{AX}$，并且使得$Y_1=\frac{1}{\sqrt{n}}\sum_i X_i$，即$Y_1=\sqrt{n}\overline{X}$，可见$Y_1\sim N(\sqrt{n}\mu,\sigma^2)$，这对应 $\mathbf{A}$ 的第一横排的元素全为$\frac{1}{\sqrt{n}}$ (某一行元素相同的正交阵是存在的：设秩为 1 的 $\mathbf{B}=\begin{pmatrix}1&\cdots&1\\ \vdots&\ddots&\vdots\\ 1&\cdots&1\end{pmatrix}$，则存在 $\mathbf{A}$ 使得，$\mathbf{A}^T\mathbf{B}\mathbf{A}=\begin{pmatrix}n&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&0\end{pmatrix}$)，此时$\sum_i Y_i^2=\mathbf{Y}^T\mathbf{Y}=\mathbf{X}^T\mathbf{A}^T\mathbf{A}\mathbf{X}=\mathbf{X}^T\mathbf{X}=\sum_i X_i^2$，则将联合密度函数用 y 来表示即为：$(\frac{1}{\sqrt{2\pi}\sigma})^n\cdot e^{-\frac{\sum_i Y_i^2-2\mu\sqrt{n}Y_1+n\mu^2}{2\sigma^2}}=(\frac{1}{\sqrt{2\pi}\sigma})^n\cdot e^{-\frac{(Y_1-\sqrt{n}\mu)^2}{2\sigma^2}}\cdot\prod_{i=2}^{n}\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{Y_i^2}{2\sigma^2}}$。——其中，作为$X_i$的线性组合的$Y_i(i\geq2)$也服从正态分布，且$Y_i(i\geq1)$两两相互独立，并且随着 n 的较大，$Y_i\dot{\sim}N(0,\sigma^2)(i\geq2)$【可证明$Y_i\sim N(0,\sigma^2)(i\geq2)$】，这点也可从用 y 表达的联合密度函数上看出。

此时$\frac{(n-1)S^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}=\frac{\sum_{i=1}^{n}[X_i^2-2X_i\overline{X}+\overline{X}^2]}{\sigma^2}=\frac{\sum_{i=1}^{n}X_i^2-n\overline{X}^2}{\sigma^2}=\frac{\sum_i Y_i^2-Y_1^2}{\sigma^2}=\frac{\sum_{i=2}^{n}Y_i^2}{\sigma^2}=\sum_{i=2}^{n}\left(\frac{Y_i-0}{\sigma}\right)^2\sim\chi^2(n-1)$。【可利用以下结论：$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$】

****

我们从这个角度作为切入点出发，也可以更简单地得到样本方差的期望：由于服从卡方分布$\chi^2(n-1)$的随机变量的期望 $E(\frac{(n-1)S^2}{\sigma^2})=n-1$，所以 $E(S^2)=\sigma^2=D[X]$。而且有 $E(\sum_{i=1}^{n}(X_i-\overline{X})^2)=(n-1)\sigma^2$；并且利用$B_2=\frac{n-1}{n}S^2$，或者$B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$还可得到 $E(B_2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}\sigma^2=\frac{n-1}{n}D[X]$；另外，利用它，还可以得到 $D(S^2)=\frac{\sigma^4}{(n-1)^2}\cdot D(\frac{(n-1)S^2}{\sigma^2})=\frac{\sigma^4}{(n-1)^2}\cdot D(\frac{(n-1)S^2}{\sigma^2})=\frac{\sigma^4}{(n-1)^2}\cdot2(n-1)=\frac{2\sigma^4}{n-1}$。因而 $D(\sum_{i=1}^{n}(X_i-\overline{X})^2)=D((n-1)S^2)=2\sigma^4(n-1)$，于是便可进一步求 $D(B_2)=D(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2)=\frac{2\sigma^4(n-1)}{n^2}$等等。——注意：该部分的$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$可以和之前的$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$进行对比，它们的期望均相等$=\sigma^2$，但后者的方差$\frac{2\sigma^4}{n}$更小<前者$S^2$的$\frac{2\sigma^4}{n-1}$，因此对于估计$\sigma^2$来说，后者更有效。

(2).$\overline{X}$与S^2相互独立。

由于$Y_1=\sqrt{n}\overline{X}$，可知$\overline{X}$仅依赖于Y_1；由$(n-1)S^2=\sum_{i=2}^{n}Y_i^2$，可知$S^2$仅依赖于$Y_i^2(i\geq2)$。而$Y_i(i\geq1)$两两相互独立。

(3).$\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$。【其中的$\overline{X}-\mu$可被$(X_{n+1}-\overline{X}_n)/\sqrt{n+1}$代替；$\overline{X}_n=\sum_{i=1}^{n}X_i$】【由于$B_2=\frac{n-1}{n}S^2$，所以有时候会这么考：$\frac{\overline{X}-\mu}{\sqrt{B_2}/\sqrt{n-1}}\sim t(n-1)$，即$\frac{\sqrt{n-1}(\overline{X}-\mu)}{\sqrt{B_2}}\sim t(n-1)$。】【有个这样的题：$\frac{(\overline{X}-\mu)^2}{(S/\sqrt{n})^2}=\frac{\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2/1}{\frac{(n-1)S^2}{\sigma^2}/(n-1)}\sim F(1,n-1)$】

证明：$\frac{\overline{X}-\mu}{S/\sqrt{n}}=\frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}}\cdot\frac{1}{\sqrt{n-1}}}\sim\frac{N(0,1)}{\chi^2(n-1)\frac{1}{\sqrt{n-1}}}=t(n-1)$。

3.定理 3：

设$X_1,X_2,\cdots,X_n$是来自总体$N(\mu_1,{\sigma_1}^2)$的一个样本，$Y_1,Y_2,\cdots,Y_n$是来自总体$N(\mu_2,{\sigma_2}^2)$的一个样本，它们之间和各自内部均相互独立，则：

(1).$\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2})$。

这是因为$\overline{X}\sim N(\mu_1,\frac{{\sigma_1}^2}{n_1})$、$\overline{Y}\sim N(\mu_2,\frac{{\sigma_2}^2}{n_2})$。

(2).$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$。

证明：$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}=\frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)}\sim F(n_1-1,n_2-1)$。

(3).当$\sigma_1^2=\sigma_2^2=\sigma^2$时，$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$。

证明：一方面，设$\overline{X}-\overline{Y}$的标准化变量为：$U=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$；另一方面，令$V=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}=\frac{(n_1-1)S_1^2}{\sigma_1^2}+\frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_1-1)+\chi^2(n_2-1)=\chi^2(n_1+n_2-2)$。于是依据 t 分布的定义：$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{U}{\sqrt{V/(n_1+n_2-2)}}\sim t(n_1+n_2-2)$。其中令$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$，可以证明$E(S_w^2)=E(\frac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{j=1}^{n_2}(Y_j-\overline{Y})^2}{n_1+n_2-2})=\sigma^2$。

【考题：还可以这么构造 U、V：$\frac{U^2}{V/(n_1+n_2-2)}\sim F(1,n_1+n_2-2)$，令其中的$n_1=n_2$、$\sigma_1^2=\sigma_2^2=\sigma^2$，则有$\frac{\left(\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{n}}}\right)^2}{\frac{(n-1)(S_1^2+S_2^2)}{\sigma^2}/(2n-2)}=\frac{\left(\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n}+\frac{1}{n}}}\right)^2}{(S_1^2+S_2^2)/2}=\frac{n(\overline{X}-\overline{Y}-(\mu_1-\mu_2))^2}{S_1^2+S_2^2}\sim F(1,2n-2)$】【似乎所有关于 t 的，平方后都为 F 了】

第七章 参数估计

在研究总体时，总体的分布或许已知，但分布中有未知参数，研究如何用样本所提供的信息来估计未知参数(的真值)问题，即是参数估计问题。参数估计问题是利用从总体抽样得到的信息，来估计总体的某些参数，或参数的某些函数。

参数估计有两种基本形式：点估计和区间估计。

7.1.点估计

在参数估计问题中，假定总体分布形式已知，未知的仅仅是一个或几个参数。利用样本函数 $g(X_1,X_2,\cdots,X_n)$——统计量，及其观测值，来给出未知参数(们)的一个估计，这就是参数的点估计。

设 $f(x,\theta)$为总体 X 的分布，其中θ指全部未知参数$(\theta_1,\theta_2,\cdots,\theta_k)$。若统计量 $g(X_1,X_2,\cdots,X_n)$的观测值 $g(x_1,x_2,\cdots,x_n)$，被视为θ的估计值，则称统计量 $g(X_1,X_2,\cdots,X_n)$为θ的一个点估计量，记为$\hat{\theta}(X_1,X_2,\cdots,X_n)$；称统计量的观测值 $g(x_1,x_2,\cdots,x_n)$为θ的一个点估计值，记为$\hat{\theta}(x_1,x_2,\cdots,x_n)$。

7.1.1 点估计的方法

1.矩估计法：

许多分布的期望与方差，都是分布的参数的函数——例如，参数为λ的泊松分布，其期望和方差均为λ；区间(a,b)上的均匀分布的期望为$\frac{a+b}{2}$，方差为$\frac{(b-a)^2}{12}$。而(总体的)期望和方差又可以分别由(总体的)一阶原点矩、二阶中心矩表示，因此许多分布的(总体的)一、二阶原点矩(二阶原点矩=二阶中心矩加上一阶原点矩的平方 $D(X)=E(X^2)-E^2(X)$)等，都是分布中未知参数的函数。

另外，根据辛钦大数定理，作为样本的$X_1,X_2,\cdots,X_n$独立同分布，则$X_1^k,X_2^k,\cdots,X_n^k$也独立同分布，具有数学期望 $E(X_i^k)=E(X^k)$。则对于任意正数ε，有$\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i^k -$

$E(X^k)| < \varepsilon\}=1$，即 X 的样本 k 阶原点矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$依概率收敛于总体 X 的 k 阶原点矩 $E(X^k)$。

而 $E(X^k)$又多是未知参数的函数，因此样本 k 阶原点矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$能较好地估计 $E(X^k)$，即未知参数的函数。——也就是说，大多数 $E(X^i)$都能写成 $g_i(\theta_1,\theta_2,\cdots,\theta_k)$，令$A_i=E(X^i)$，则$A_i=E(X^i)=g_i(\theta_1,\theta_2,\cdots,\theta_k)$，$i=1,2,\cdots,k$。【i 不一定从 1 取到 k，因为可能 $E(X^1)$中不含有$(\theta_1,\theta_2,\cdots,\theta_k)$中的任何一个，此时要接着求 $E(X^2)\cdots$，直到一共有 k 个 $E(X^r)$的表达式中含有θ】

解$A_i=\frac{1}{n}\sum_{j=1}^{n}X_j^i=g_i(\theta_1,\theta_2,\cdots,\theta_k)$，$i=1,2,\cdots,k$ 这个 k 阶方程组，会得到 $\hat{\theta}_i=\hat{\theta}_i(X_1,X_2,\cdots,X_n)$，$i=1,2,\cdots,k$，这称为$\theta_1,\theta_2,\cdots,\theta_k$的矩估计量。$\hat{\theta}_i(x_1,x_2,\cdots,x_n)$称为$\theta_1,\theta_2,\cdots,\theta_k$的矩估计值。——这种利用样本原点矩得到点估计的方法就称为矩估计法。

【正如以辛钦大数定理为依据，A_k能较好地估计 $E(X^k)$一样，仍用辛钦大数定理，可证明样本 k 阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$能较好地估计总体 X 的 k 阶中心矩 $E([X-E(X)]^k)$【而不是 $E((X_i-\overline{X})^k)$】——并不是$(X_i-\overline{X})^k$也独立同分布，且并没有 $E((X_i-\overline{X})^k)=E([X-E(X)]^k)$——而是比如根据$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$，有 $\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2$，即$B_2=A_2-A_1^2$，再根据依概率收敛的性质，$B_2$依概率收敛于$A_2-A_1^2$所依概率收敛于的 $E(X^2)-E^2(X)$，即 $D(X)$，即 $E([X-E(X)]^2)$。所以也可用B_k来估计/代替 $E([X-E(X)]^k)$，而 $E([X-E(X)]^k)$又是各阶原点矩 $E(X^r)(r\le k)$的函数，(比如 $E([X-E(X)]^2)=B_2=A_2-A_1^2$)，因此 $E([X-E(X)]^k)$也多是未知参数的函数，因此可以有$B_i=E([X-E(X)]^i)=g_i(\theta_1,\theta_2,\cdots,\theta_k)$，$i=1,2,\cdots,k$。并以此来列方程(组)。这种方法本质上也是矩估计法。】

由于总体矩在用相应样本矩代替时，对于所采用的矩的阶数有一定的随意性，所以矩估计法得到的矩估计量的答案不唯一，因此考试一般不考。

****

(1).已知总体的分布律为 $P\{X=0\}=\theta^2$；$P\{X=1\}=2\theta(1-\theta)$；$P\{X=2\}=\theta^2$；$P\{X=3\}=1-2\theta$；$0<\theta<0.5$，未知。利用总体的样本值：3,1,3,0,3,1,2,3。——求θ的矩估计量和矩估计值。

$E(X)=3\text{-}4\theta=\overline{x}=2$，于是矩估计量$\hat{\theta}=\frac{3-\overline{X}}{4}$、矩估计值$\hat{\theta}=\frac{1}{4}$。

(2).设 X 的概率密度函数为 $f(x)=(1+\theta)x^{\theta}$，$0<x<1$；0，else。其中$\theta>-1$，$X_1,X_2,\cdots,X_n$是来自总体的一个样本，$x_1,x_2,\cdots,x_n$为样本的样本值——求θ的矩估计。

该分布中只有一个未知参数，因此只需要求 X 的一阶矩，即 $E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)\cdot dx=\int_0^1(1+\theta)x^{1+\theta}\cdot dx=\frac{1+\theta}{2+\theta}=A_1=\overline{X}$，得到$\theta$的矩估计量$\hat{\theta}=\frac{2\overline{X}-1}{1-\overline{X}}$，矩估计值$\hat{\theta}=\frac{2\overline{x}-1}{1-\overline{x}}$。

(3).设$X_1,X_2,\cdots,X_n$是取自总体 X~B(m,p)的一个样本。求 p 的矩估计量。

$E(X)=np=\overline{X}$，$E(X^2)=D(X)+E^2(X)=np(1-p)+(kp)^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$，可解出$\hat{p}=\frac{\overline{X}-S^2}{\overline{X}}$，$\hat{k}=[\frac{\overline{X}^2}{\overline{X}-S^2}]$，该方括号表示取整。

2.最大似然估计法：(MLE：maximum likely estimate)

依据小概率原理：(1).概率很小的事件在一次实验中几乎是不出现的；(2).概率很小(=p)的事件在次数很多的重复实验中几乎是一定会出现的，$\lim_{n\to\infty}1-(1-p)^n=1$。——的第一点，我们认为已然出现的样本值$x_1,x_2,\cdots,x_n$，在横向比较时，最有可能是在所有可能的样本值的组合中，出现几率最大的；在纵向即自身比较时(其实等同于横向，不同 x 组之间与 X 组的不同取法是等效的)，参数θ的取值理应使得已然出现的$x_1,x_2,\cdots,x_n$出现概率最大(已知结果，去拟合原因：让它的出现合法化、合理，正如之所以所有参数都这么巧合，使得规则允许宇宙繁衍，宇宙允许智慧出现，被解释为如果不是这样就没有人正在回答这个问题一样)。

设 f(x;θ)为总体 X 的分布，其中θ指全部未知参数$(\theta_1,\theta_2,\cdots,\theta_k)$。$X_1,X_2,\cdots,X_n$是来自总体的一个样本，$x_1,x_2,\cdots,x_n$为样本的观测值，把样本值$x_1,x_2,\cdots,x_n$出现的概率称为样本似然函数，记作 $L(\theta)=P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}=P\{X_1=x_1\}P\{X_2=x_2\}\cdots P\{X_n=x_n\}$或$\approx f(x_1,x_2,\cdots,x_n;\theta)\cdot\vartriangle x_1\cdot\vartriangle x_2\cdots\vartriangle x_n=f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)\cdot\prod_{i=1}^{n}\vartriangle x_i$。由于$\prod_{i=1}^{n}\vartriangle x_i$与$\theta$无关，$L(\theta)$与 $f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$同最大值点，因此设连续型随机变量的 $L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)$。使得 $L(\theta)$达到最大值的θ的取值$\hat{\theta}(x_1,x_2,\cdots,x_n)$叫做$\theta$的最大似然估计值，$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$称为$\theta$的最大似然估计量。

我们可以用微积分中求极大值的方法来求 $L(\theta)$的最大值点，但可看出 $L(\theta)$多为多个函数之积，直接求导很复杂(参考两个函数之积求导)；而由于对数函数单调递增，$\ln L(\theta)$与 $L(\theta)$有相同的最大值点，因此我们可转而求 $L(\theta)$的导数=0 的点——但有时候θ为$(\theta_1,\theta_2,\cdots,\theta_k)$，此时要求 $L(\theta)$对每个θ_i的偏导数；并且有时候有些导数和偏导≠0，此时需要综合判断。

性质：$\hat{\theta}$为θ的最大似然估计量，若 g(x)单调，则 $g(\hat{\theta})$为 $g(\theta)$的最大似然估计量。

****

续(1).已知总体的分布律为 $P\{X=0\}=\theta^2$；$P\{X=1\}=2\theta(1-\theta)$；$P\{X=2\}=\theta^2$；$P\{X=3\}=1-2\theta$；$0<\theta<0.5$，未知。利用总体的样本值：3,1,3,0,3,1,2,3。——求θ的最大似然估计值。

$L(\theta)=(1-2\theta)^4(2\theta(1-\theta))^2\theta^2\theta^2$，$\frac{d\ln L(\theta)}{d\theta}=\frac{d[4\ln(1-2\theta)+\ln 4+6\ln\theta+2\ln(1-\theta)]}{d\theta}=\frac{-4}{1-2\theta}+\frac{3}{\theta}-\frac{1}{1-\theta}=0$，于是$-4\theta(1-\theta)+3(1-2\theta)(1-\theta)-\theta(1-2\theta)=0$，得到$12\theta^2-14\theta+3=0$，于是$\theta=\frac{14\pm\sqrt{14^2-4\cdot12\cdot3}}{2\cdot12}$，而 $0<\theta<0.5$，所以最大似然估计值$\hat{\theta}=\frac{7-\sqrt{13}}{12}$。

类(2).设总体 X 的概率密度函数为 $f(x)=\theta\cdot x^{\theta-1}$，$0<x<1$；0，else。其中$\theta>0$。求$\theta$的最大似然估计量。

于是 $L(\theta)=\theta^n\cdot(x_1\cdot x_2\cdots\cdot x_n)^{\theta-1}$，$\ln L(\theta)=n\cdot\ln(\theta)+(\theta-1)\sum_{i=1}^n\ln x_i$，$\frac{d\ln L(\theta)}{d\theta}=\frac{n}{\theta}+\sum_{i=1}^n\ln x_i=0$，于是$\hat{\theta}=-\frac{n}{\sum_{i=1}^n\ln x_i}$，这里有个负号并不要紧，因为 $0<x<1$ 所对应的$\ln x_i$均是负的，因而算出来的参数θ的最大似然估计量$\hat{\theta}$的值是正的。

(3).设$X_1,X_2,\cdots,X_n$是取自总体 $X\sim B(1,p)$的一个样本。求 p 的最大似然估计量。

对于 0-1 分布，其分布律为 $P\{X=x\}=p^x(1-p)^{1-x}$，其中 x=0 或 1。于是$L(p)=\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$，于是$\ln L(p)=\ln(p)\cdot\sum_{i=1}^n x_i+(n-\sum_{i=1}^n x_i)\cdot\ln(1-p)$，于是$\frac{d\ln L(p)}{dp}=\frac{1}{p}\sum_{i=1}^n x_i-\frac{1}{1-p}(n-\sum_{i=1}^n x_i)=0$。得到$(\frac{1}{p}+\frac{1}{1-p})\sum_{i=1}^n x_i=\frac{n}{1-p}$，即$(\frac{1-p}{p}+1)\sum_{i=1}^n x_i=n$，即有$\frac{1}{p}\cdot\frac{\sum_{i=1}^n x_i}{n}=1$，即有$\hat{p}=\overline{X}$。

——设$X_1,X_2,\cdots,X_n$是取自总体 $X\sim B(m,p)$的一个样本。求 p 的最大似然估计量。

对于二项分布，其分布律为 $P\{X=x\}=C_m^x\cdot p^x(1-p)^{m-x}$。于是$L(p)=\prod_{i=1}^n p^{x_i}(1-p)^{m-x_i}\prod_{i=1}^n C_m^{x_i}=p^{\sum_{i=1}^n x_i}(1-p)^{nm-\sum_{i=1}^n x_i}\prod_{i=1}^n C_m^{x_i}$，于是$\ln L(p)=\ln(p)\cdot\sum_{i=1}^n x_i+(nm-\sum_{i=1}^n x_i)\cdot\ln(1-p)+\ln(\prod_{i=1}^n C_m^{x_i})$，于是$\frac{d\ln L(p)}{dp}=\frac{1}{p}\sum_{i=1}^n x_i-\frac{1}{1-p}(nm-\sum_{i=1}^n x_i)=0$。得到$(\frac{1}{p}+\frac{1}{1-p})\sum_{i=1}^n x_i=\frac{nm}{1-p}$，即$(\frac{1-p}{p}+1)\sum_{i=1}^n x_i=nm$，即有$\frac{1}{p}\cdot\frac{\sum_{i=1}^n x_i}{nm}=1$，即有$\hat{p}=\frac{\overline{X}}{m}$。

(4).设总体 $X\sim E(\lambda)$，即 $f(x)=\lambda\cdot e^{-\lambda x}$，$x>0$；$=0$，$x\le0$。求$\lambda$的最大似然估计量。

$\ln L(\lambda)=n\cdot\ln(\lambda)-\lambda\sum_{i=1}^n x_i$，于是$\frac{d\ln L(\lambda)}{d\lambda}=\frac{n}{\lambda}-\sum_{i=1}^n x_i=0$，于是$\hat{\lambda}=\frac{1}{\overline{X}}$。——根据指数分布的 $E(X)=\frac{1}{\lambda}=A_1=\overline{X}$，可见矩估计量也有同样的表达式。

多个θ型(5).$L(\theta,\mu)=\prod_{i=1}^n\frac{1}{\theta}e^{-(x_i-\mu)/\theta}$，其中$x_i\ge\mu$，$\theta>0$，则 $\ln L(\theta,\mu)=-n\cdot\ln\theta-\frac{1}{\theta}\cdot\sum_{i=1}^n(x_i-\mu)$，于是$\frac{\partial\ln L(\theta,\mu)}{\partial\theta}=-\frac{n}{\theta}+\frac{1}{\theta^2}\cdot\sum_{i=1}^n(x_i-\mu)=0$，$\frac{\partial\ln L(\theta,\mu)}{\partial\mu}=\frac{n}{\theta}>0$，后者示意在$\mu\le\min x_i$的区间上，$\ln L(\theta,\mu)$关于$\mu$为增函数，因此$\hat{\mu}=\min x_i$，将其代入$-\frac{n}{\theta}+\frac{1}{\theta^2}\cdot$

$\sum_{i=1}^{n}(x_i-\mu)=0$，即可得到$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})=\frac{1}{n}\sum_{i=1}^{n}(x_i-\min x_i)$，这俩便分别是$\theta$和$\mu$的最大似然估计值。

(6).设 X 在[a,b]上服从均匀分布，其中 a,b(a<b)为未知参数，求 a,b 的最大似然估计：

$L(a,b)=\left(\frac{1}{b-a}\right)^n$，$\ln L(a,b)=-n\cdot\ln(b-a)$，$\frac{\partial \ln L(a,b)}{\partial a}=\frac{n}{b-a}>0$，$\frac{\partial \ln L(a,b)}{\partial b}=-\frac{n}{b-a}<0$，又因 $a<x_i<b$，因此 $a<\min x_i$，$b>\max x_i$。而 $L(a,b)$分别关于a、b单增和单减，因此$\hat{a}=\min x_i$、$\hat{b}=\max x_i$，分别是 a 和 b 的最大似然估计值。

7.1.2 点估计量的评价准则

从前面一节可以看到，同一参数可以有不同的估计量。比如θ的矩估计量本身就有很多个，再加上θ的最大似然估计量，可见许多统计量均可作为未知参数的估计量。于是就存在判断同一参数的不同估计量之间孰优孰劣的问题。即不同估计法的比较问题。

数理统计中常见的估计量评价准则有无偏性、有效性、相合性(一致性)。

1.无偏性

设$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$为未知参数$\theta$的一个估计量，若 $E(\hat{\theta})$存在，且 $E(\hat{\theta})=\theta$，则称$\hat{\theta}$为θ的无偏估计量。也称$\hat{\theta}$作为θ的估计具有无偏性。否则称$\hat{\theta}$为θ的有偏估计量。

在实际应用中，$E(\hat{\theta})-\theta$称为系统误差，无偏的含义就是无系统误差，即 $E(\hat{\theta})-\theta=0$。

若 $E(\hat{\theta})$存在，且$\lim\limits_{n\to\infty}E(\hat{\theta})=\theta$，则称$\hat{\theta}$为$\theta$的渐进无偏估计量。

****

(1).设总体 X 的期望μ与方差σ^2均存在，可以证明，样本均值$\overline{X}$、样本方差S^2，分别是的μ、σ^2的无偏估计量；且$S_n^2=B_2=\frac{n-1}{n}S^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$是$\sigma^2$的渐进无偏估计量。

这是因为由于样本的独立同分布和数学期望的性质，$E(\overline{X})=\mu$，$E(S^2)=\sigma^2$[我们之前在数字特征处证明过它，利用$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$。而 $E(S_n^2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}\sigma^2\neq\sigma^2$，因此$S_n^2$不是总体方差$\sigma^2$的无偏估计量；但因$\lim\limits_{n\to\infty}E(S_n^2)=\sigma^2$，所以$S_n^2$是总体方差$\sigma^2$的渐进无偏估计量。

(2).续上一节的(6).，当其中 a=0 时，证明$\frac{n+1}{n}\hat{b}$是 b 的无偏估计量。

由总体 X 的分布函数 F(x)=0，x<0；$\frac{x}{b}$，0≤x≤b；1，x>b。那么$\hat{b}=\hat{b}(X_1,X_2,\cdots,X_n)$的分布函数为：$F_{\hat{b}}(x)=F(x,x,\cdots,x)=[F(x)]^n=\frac{x^n}{b^n}$，0≤x≤b；于是$\hat{b}$的密度函数$f_{\hat{b}}(x)=\hat{b}(x,x,\cdots,x)=\frac{n\cdot x^{n-1}}{b^n}$，0≤x≤b。因此 $E(\hat{b})=\int_{-\infty}^{+\infty}x\cdot f_{\hat{b}}(x)\cdot dx=\int_0^b x\cdot\frac{n\cdot x^{n-1}}{b^n}\cdot dx=\frac{n}{n+1}b$。【利用了 3.5 中多个随机变量相互独立，以及 3.6 中的$f_{\hat{b}}(x)=\frac{dF_{\hat{b}}(x)}{dx}$】

因此 $E(\frac{n+1}{n}\cdot\hat{b})=b$。【注：$F_{\hat{b}}(x)$不是通常意义上的$F_{\hat{b}}(\hat{b})$，即不像$F_Z(z)$一样，而是像将 $F(x_1,x_2,\cdots,x_n)=F_{X_1}(x_1)\cdot F_{X_2}(x_2)\cdots F_{X_n}(x_n)$中的$x_i=x$后所得的分布函数；且这里的$\hat{b}(X_1,X_2,\cdots,X_n)$相当于 g(X,Y)而不是 f(x,y)】

2.有效性

有效性只在同一参数的不同无偏估计量之间比较。用于评价它们的优劣。

设$\hat{\theta}_1,\hat{\theta}_2$是参数$\theta$的两个无偏估计量，因而取值都在$\theta$附近波动。在样本容量相同的条件下，若$\hat{\theta}_1$在$\theta$附近取值的概率比$\hat{\theta}_2$更大，或者说$\hat{\theta}_1$的观测值比$\hat{\theta}_2$的更密集地在$\theta$附近，那么认为用$\hat{\theta}_1$估计$\theta$比用$\hat{\theta}_2$更好一些。因此提出了无偏估计量的有效性准则。

若$\hat{\theta}_1(X_1,X_2,\cdots,X_n)$、$\hat{\theta}_2(X_1,X_2,\cdots,X_n)$均为无偏估计量，且 $D(\hat{\theta}_1)<D(\hat{\theta}_2)$，即 $E[(\hat{\theta}_1-\theta)^2]<E[(\hat{\theta}_2-\theta)^2]$，则称$\hat{\theta}_1$比$\hat{\theta}_2$有效。

****

比如 $E(\overline{X})=E(X_i)=\theta=\mu$，$D(\overline{X})=\frac{\sigma^2}{n}<\sigma^2=D(X_i)$，因此两个对于$\mu$的无偏估计量$\overline{X}$、$X_i$，$\overline{X}$相比于$X_i$，在估计同一参数$\mu$时，更有效。

3.相合性(一致性)

不要求估计量为无偏估计量；无偏性和有效性均是在样本容量有限的前提下讨论的。6.1.4 经验分布函数处的格利文科定理指出，样本容量越大，经验分布函数越逼近总体分布函数。因此我们希望估计量的观测值，随着样本容量的增大，也越来越逼近未知参数。因此在大样本的前提下，提出了估计量的相合性准则。

设$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$为未知参数$\theta$的一个估计量，若对于任意正数$\varepsilon$，有$\lim_{n\to\infty}P\{|\hat{\theta}-\theta|<\varepsilon\}=1$或者$\lim_{n\to\infty}P\{|\hat{\theta}-\theta|\geq\varepsilon\}=0$，则称$\hat{\theta}$为$\theta$的相合估计量。【与渐进无偏估计量有区别】

相合性是对估计量的一个基本要求，不具备相合性的估计量是不予考虑的。由辛钦大数定理(独立同分布，(+有同期望)即可)可知，由于$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$依概率收敛于$E(X^k)$，所以样本 k 阶原点矩$A_k$是总体 X 的 k 阶原点矩 $E(X^k)$的相合估计量。——根据这个思想，会有一系列连锁反应：对于一个恰好有 D(λ)=E(λ)=的总体 X~P(λ)，它的样本的二阶中心矩B_2，将有 $E(B_2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}E(S^2)=\frac{n-1}{n}D(X_i)=\frac{n-1}{n}\cdot\lambda$，根据定义$\lim\limits_{n\to\infty}E(B_2)=\lim\limits_{n\to\infty}\frac{n-1}{n}\cdot\lambda=\lambda$，$B_2$是λ的渐进无偏估计量；又因泊松分布的总体的一、二阶原点矩 $E(X)$、$E(X^2)$存在，而根据大数定律，其样本的一、二阶原点矩$\frac{1}{n}\sum_{i=1}^{n}X_i$、$\frac{1}{n}\sum_{i=1}^{n}X_i^2$又依概率收敛于总体的一、二阶原点矩 $E(X)$、$E(X^2)$，而样本的二阶中心矩B_2又是A_1、A_2的函数：根据$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$，有$\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2$，即$B_2=A_2-A_1^2$依概率收敛于 $E(X^2)-E^2(X)=D(X)$，因此B_2也依概率收敛，且依概率收敛于总体的二阶中心矩 $D(X)$，而不是 $E(B_2)$。——而此时 $D(X)=\lambda$，即B_2依概率收敛于λ，即B_2是λ的一致(相合)估计量。——且进一步地，S^2也是λ的一致(相合)估计量。

可见B_2是λ的渐进无偏估计量并不能直接推出B_2是λ的一致(相合)估计量，因为并没有B_2依概率收敛于 $E(B_2)$，而是总体的二阶中心矩 $D(X)$；看样子这便是为什么书上没有这样介绍的原因：其因果关系不强，并非一定有一个能推出另一个；链接到 7.1.1 点估计的方法中的 1.矩估计法。

****

(1).续无偏性中的(2).，当 a=0 时，证明$\frac{n+1}{n}\hat{b}$是 b 的相合估计量。

$E(\hat{b}^2)=\int_{-\infty}^{+\infty}x^2\cdot f_{\hat{b}}(x)\cdot dx=\int_0^b x^2\cdot\frac{n\cdot x^{n-1}}{b^n}\cdot dx=\frac{n}{n+2}b^2$。$D(\hat{b})=\frac{n}{n+2}b^2-\frac{n^2}{(n+1)^2}b^2=\frac{n(n+1)^2-n^2(n+2)}{(n+2)(n+1)^2}b^2=\frac{n}{(n+2)(n+1)^2}b^2$。根据切比雪夫不等式以及概率的性质，像方差的性质(4).的证明一样，我们有 $0\le P\{|X-E(X)|\ge\varepsilon\}\le\frac{D(X)}{\varepsilon^2}$，即 $0\le P\{|\frac{n+1}{n}\cdot\hat{b}-E(\frac{n+1}{n}\cdot\hat{b})|\ge\varepsilon\}\le\frac{D(\frac{n+1}{n}\cdot\hat{b})}{\varepsilon^2}$，利用 $E(\frac{n+1}{n}\cdot\hat{b})=b$，以及 $D(\hat{b})=\frac{n}{(n+2)(n+1)^2}b^2$，我们有 $0\le P\{|\frac{n+1}{n}\cdot\hat{b}-b|\ge\varepsilon\}\le\frac{\frac{(n+1)^2}{n^2}\frac{n}{(n+2)(n+1)^2}b^2}{\varepsilon^2}=\frac{b^2}{(n+2)n\varepsilon^2}$，于是$\lim\limits_{n\to\infty}P\{|\frac{n+1}{n}\cdot\hat{b}-b|\ge\varepsilon\}=0$，因此$\frac{n+1}{n}\hat{b}$是 b 的相合估计量。

7.2 区间估计

7.2.1 置信区间

我们希望求出一个尽可能小的区间$(\hat{\theta}_1,\hat{\theta}_2)$，使得 P$\{\hat{\theta}_1<\theta<\hat{\theta}_2\}=1-\alpha$，较大。我们称$1-\alpha$为置信概率、置信度 or 置信水平，其中$\alpha$为之前介绍的上$\alpha$分位数。并称随机区间$(\hat{\theta}_1,\hat{\theta}_2)$是$\theta$的，置信度为$1-\alpha$的，(双侧)置信区间；称$\hat{\theta}_1$和$\hat{\theta}_2$是该双侧置信区间的置信下限和置信上限。

对θ作区间估计，就是要设法(构造统计量)找出两个只依赖于样本的界限，使得θ以很大可能被包含在$(\hat{\theta}_1,\hat{\theta}_2)$内，即要使 P$\{\hat{\theta}_1<\theta<\hat{\theta}_2\}$尽可能大，即可靠度高；并且估计的精度要尽可能的高，即使得区间长度$\hat{\theta}_2-\hat{\theta}_1$尽可能短(精度也和 D(X)、$\sigma^2$相关联)。——但是，可靠度和精度是一对矛盾，我们通常在保证可靠度的条件下，尽可能地提高精度。

寻找置信区间的方法，一般从确定误差限入手，即找到一个$\delta(\alpha)$，使得 P$\{|\hat{\theta}-\theta|<\delta\}=1-\alpha$，此时 P$\{\hat{\theta}-\delta<\theta<\hat{\theta}+\delta\}=1-\alpha$中的$\hat{\theta}-\delta$便相当于$\hat{\theta}_1$，$\hat{\theta}+\delta$便相当于$\hat{\theta}_2$。

寻求置信区间的一般方法(单侧类似)：

(1).明确问题，求什么参数($\hat{\theta}_1$还是$\hat{\theta}_2$)的置信区间？置信水平$1-\alpha$是多少？

(2).寻找参数θ的一个良好的点估计$\hat{\theta}(X_1,X_2,\cdots,X_n)$，用于提示自己接下来选什么长相的 S(限定一下可能的选择范围)。

(3).寻找一个待估参数θ和其估计量的函数(叫做枢轴量)S$(\hat{\theta},\theta)$，要求其分布函数已知(因此要在服从已知的四大分布的来自正态总体的常用统计量中选择)，并且不含除了θ外的任何未知参数(比如含$\hat{\theta}_1$，而不含$\hat{\theta}_2$等)。

(4).根据 S$(\hat{\theta},\theta)$的分布，确定$\delta(\alpha)$的前身——S$(\hat{\theta},\theta)$的上$f(\alpha)=\frac{\alpha}{2}$分位数，使得 P$\{|S(\hat{\theta},\theta)|<S_{\frac{\alpha}{2}}(\hat{\theta},\theta)\}$=给定的置信水平$1-\alpha$。——不过对于有些非对称的 S$(\hat{\theta},\theta)$分布，我们得根据 S$(\hat{\theta},\theta)$的分布，确定 P$\{S_{f_1(\alpha)}(\hat{\theta},\theta)<S(\hat{\theta},\theta)<S_{f_2(\alpha)}(\hat{\theta},\theta)\}$= P$\{S(\hat{\theta},\theta)>S_{f_1(\alpha)}(\hat{\theta},\theta)\}$−P$\{S(\hat{\theta},\theta)>S_{f_2(\alpha)}(\hat{\theta},\theta)\}=f_1(\alpha)-f_2(\alpha)$=给定的$1-\alpha$，中的$f_1(\alpha)$、$f_2(\alpha)$，再变形化为类似于 P$\{\hat{\theta}-\delta<\theta<\hat{\theta}+\delta\}=1-\alpha$或者 P$\{\hat{\theta}_1<\theta<\hat{\theta}_2\}=1-$

α的式子。【在 2.4.2 随机变量的分布函数性质中，利用 P(A-B)=P(A)-P(AB)，可以得到 $P\{a<X\le b\}=P\{\{X>a\}\cap\{X\le b\}\}=P\{\{X>a\}-\{X>b\}\}=P\{X>a\}-P\{\{X>a\}\cap\{X>b\}\}=P\{X>a\}-P\{X>b\}$】

****

(一).求一些典型的被估参数(一般是μ、σ^2)的双侧置信区间

(1).假设我们要估计一个正态总体 X 的分布(密度)函数中的μ值，若已知σ^2：

首先，选取μ的点估计为其矩估计量$\overline{X}$；在来自正态总体的 9 个常用统计量中寻找一个含$\overline{X}$和被估计参数μ的估计量，由于σ^2已知，所以其中可以含有σ^2。——一个合适的选择便是定理 1 的(2).$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，你可能想用定理 2 的(3).$\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，因为其中没有$\sigma^2$这个$\mu$之外的参数岂不是更好，不过你需要知道，$S$像$\overline{X}$一样，是随机变量，而对$\mu$的估计这个既有目的，将允许$\overline{X}$会被抽样调查到，而$S$不会——一方面已经知道了$\sigma^2$，不需要再统计和计算出$S$；另一方面，我们的目的是求$\mu$，对抽到的样本的$S$并不关心。

于是我们便已经找到了符合要求的 $S(\hat{\theta},\theta)=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$，接下来便是根据$1-\alpha$，得到$f_1(\alpha)-f_2(\alpha)=1-\alpha$，一方面我们可选用诸如$f_1(\alpha)=1-\alpha$、$f_2(\alpha)=0$，对应$S_{f_1(\alpha)}(\hat{\theta},\theta)=z_{1-\alpha}=-z_\alpha$和$S_{f_2(\alpha)}(\hat{\theta},\theta)=z_0$，但是这样的中间步骤所得区间是半无限长的$(-z_\alpha,z_0)=(-z_\alpha,+\infty)$，接下来得到的置信区间也将是半无限大的，其精度非常低；因此，对于这种对称的 $S(\hat{\theta},\theta)$，我们倾向于选择$f_1(\alpha)=1-\frac{\alpha}{2}$、$f_2(\alpha)=\frac{\alpha}{2}$，使得$z_{1-\alpha}=-z_\alpha$，可以证明该区间在保证其所对应的面积$=1-\alpha$的条件下长度最短——该区间左移右移时，为保证面积$=1-\alpha$，区间长度会增大，精度降低。此时的中间步骤所得区间为$(z_{1-\frac{\alpha}{2}},z_{\frac{\alpha}{2}})$，即$(-z_{\frac{\alpha}{2}},z_{\frac{\alpha}{2}})$，即有 $P\{|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}|<z_{\frac{\alpha}{2}}\}=1-\alpha$，得到 $P\{\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_{\frac{\alpha}{2}}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}\cdot z_{\frac{\alpha}{2}}\}=1-\alpha$。于是最终得到$\mu$的置信度为$1-\alpha$的双侧置信区间$(\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_{\frac{\alpha}{2}},\overline{X}+\frac{\sigma}{\sqrt{n}}\cdot z_{\frac{\alpha}{2}})$。

(2).同样对于单个正态总体，同样要估计其μ值，不过现在未知σ^2：

现在才该考虑使用含有μ的估计量$\overline{X}$的 9 个常用统计量中的定理 2 的(3).$\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，因其含被估参数$\mu$及其估计量$\overline{X}$，而不含其他未知参数，如$\sigma^2$(当然，此时题目应当告诉你样本的$S$)。——对于 $S(\hat{\theta},\theta)=\frac{\overline{X}-\mu}{S/\sqrt{n}}$，同样选取$f_1(\alpha)=1-\frac{\alpha}{2}$、$f_2(\alpha)=\frac{\alpha}{2}$，使得$f_1(\alpha)-f_2(\alpha)$=给定的$1-\alpha$，此时$S_{f_1(\alpha)}(\hat{\theta},\theta)=t_{1-\frac{\alpha}{2}}(n-1)=-t_{\frac{\alpha}{2}}(n-$

1)、$S_{f_2(\alpha)}(\hat{\theta},\theta)=t_{\frac{\alpha}{2}}(n-1)$。于是 $P\{|\frac{\overline{X}-\mu}{S/\sqrt{n}}|<t_{\frac{\alpha}{2}}(n-1)\}=1-\alpha$。因此$(\overline{X}-\frac{S}{\sqrt{n}}\cdot t_{\frac{\alpha}{2}}(n-1),\overline{X}+\frac{S}{\sqrt{n}}\cdot t_{\frac{\alpha}{2}}(n-1))$。

【有时候当题目暗示 n 很大，或者明示$n-1>45$时，此时 $t(n-1)$近似地服从标准正态分布 $N(0,1)$，即有$\frac{\overline{X}-\mu}{S/\sqrt{n}}\dot{\sim}N(0,1)$，此时相当于将(1).中的$\sigma$用$S$代替；或者说用$z_{\frac{\alpha}{2}}$替代(2).中的$t_{\frac{\alpha}{2}}(n-1)$。即有 $P\{\overline{X}-\frac{S}{\sqrt{n}}\cdot z_{\frac{\alpha}{2}}<\mu<\overline{X}+\frac{S}{\sqrt{n}}\cdot z_{\frac{\alpha}{2}}\}=1-\alpha$】【另外，$\frac{S}{\sqrt{n}}\cdot t_{\frac{\alpha}{2}}(n-1)$这个东西就是大学物理实验中的等精度多次测量的 A 类不确定度表达式 (的来源)】

(3).对于单个正态总体，待估参数为σ^2，现已知μ：

利用定理 1 的(3).$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$，$P\{\chi^2_{1-\frac{\alpha}{2}}(n)<\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2<\chi^2_{\frac{\alpha}{2}}(n)\}=1-\alpha$。于是$\sigma^2$的置信度为$1-\alpha$的双侧置信区间为$(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)},\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)})$。【卡方分布、F 分布的概率密度函数不是对称的，但习惯上仍然选取对称的百分位点(上分位数)来计算未知参数的置信区间(这些密度函数在 n 不同时，形状也变了，因此其长度最小的区间的两个端点的最优解因 n 而异，且每个都不好求。)】

【为什么(1).中体现出来了μ的矩估计量$\overline{X}$，而这里没有体现出来σ^2的无偏估计量S^2呢(σ^2的矩估计量为$B_2=\frac{n-1}{n}S^2$，却只是σ^2的渐进无偏估计量，之前提及过)？——其实是体现出来了的：$(n-1)S^2=\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$——但它们的不完全吻合和下一个中的真正体现，说明并非一定需要步骤(2).——另外，为什么这里不选择(1).的$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$作为枢轴量呢？它也没有体现$\sigma^2$的良好点估计$S^2$，但$\overline{X}$比$\sum_{i=1}^{n}(X_i-\mu)^2$怕是好多了吧】

(4).对于单个正态总体，待估参数为σ^2，未知μ：

利用含有σ^2和S^2而不含有μ的，定理 2 的(1).$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，可得 $P\{\chi^2_{1-\frac{\alpha}{2}}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\frac{\alpha}{2}}(n-1)\}=1-\alpha$。于是$\sigma^2$的置信度为$1-\alpha$的双侧置信区间为$(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)})$。

(5).对于两个正态总体，待估参数为$\mu_1-\mu_2$，已知${\sigma_1}^2,{\sigma_2}^2$：

利用定理 3 的(1).$\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2})$，对应的$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}}\sim N(0,1)$，可得 $P\{z_{1-\frac{\alpha}{2}}<\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}}<z_{\frac{\alpha}{2}}\}=1-\alpha$。于是$\mu_1-\mu_2$的置信度为$1-\alpha$的双侧置信区间为$(\overline{X}-\overline{Y}-z_{\frac{\alpha}{2}}\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}},\overline{X}-\overline{Y}+z_{\frac{\alpha}{2}}\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}})$。

(6).对于两个正态总体，待估参数为$\mu_1-\mu_2$，未知$\sigma_1{}^2,\sigma_2{}^2$，但$\sigma_1^2=\sigma_2^2=\sigma^2$：

利用定理 3 的(3).$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$，可得 $P\{t_{1-\frac{\alpha}{2}}(n_1+n_2-2)<\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}<t_{\frac{\alpha}{2}}(n_1+n_2-2)\}=1-\alpha$。于是$\mu_1-\mu_2$的置信度为$1-\alpha$的双侧置信区间为$(\overline{X}-\overline{Y}-t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\overline{X}-\overline{Y}+t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$。

(7).对于两个正态总体，待估参数为$\frac{\sigma_1{}^2}{\sigma_2{}^2}$，已知$\mu_1,\mu_2$：

利用定理 1 的(3).$\frac{\sum_{i=1}^n\left(\frac{X_i-\mu_1}{\sigma_1}\right)^2/n_1}{\sum_{i=1}^n\left(\frac{Y_i-\mu_2}{\sigma_2}\right)^2/n_2}\sim F(n_1,n_2)$，可得 $P\{F_{1-\frac{\alpha}{2}}(n_1,n_2)<\frac{\sum_{i=1}^n(X_i-\mu_1)^2/\sigma_1^2 n_1}{\sum_{i=1}^n(Y_i-\mu_2)^2/\sigma_2^2 n_2}<F_{\frac{\alpha}{2}}(n_1,n_2)\}=1-\alpha$，即 $P\{F_{1-\frac{\alpha}{2}}(n_1,n_2)\frac{\sigma_1^2 n_1}{\sigma_2^2 n_2}<\frac{\sum_{i=1}^n(X_i-\mu_1)^2}{\sum_{i=1}^n(Y_i-\mu_2)^2}<F_{\frac{\alpha}{2}}(n_1,n_2)\frac{\sigma_1^2 n_1}{\sigma_2^2 n_2}\}=1-\alpha$，于是$\frac{\sigma_1{}^2}{\sigma_2{}^2}$的置信度为$1-\alpha$的双侧置信区间为
$(\frac{1}{F_{\frac{\alpha}{2}}(n_1,n_2)}\frac{n_2}{n_1}\frac{\sum_{i=1}^n(X_i-\mu_1)^2}{\sum_{i=1}^n(Y_i-\mu_2)^2},\frac{1}{F_{1-\frac{\alpha}{2}}(n_1,n_2)}\frac{n_2}{n_1}\frac{\sum_{i=1}^n(X_i-\mu_1)^2}{\sum_{i=1}^n(Y_i-\mu_2)^2})$，即
$(F_{1-\frac{\alpha}{2}}(n_2,n_1)\frac{n_2}{n_1}\frac{\sum_{i=1}^n(X_i-\mu_1)^2}{\sum_{i=1}^n(Y_i-\mu_2)^2},F_{\frac{\alpha}{2}}(n_2,n_1)\frac{n_2}{n_1}\frac{\sum_{i=1}^n(X_i-\mu_1)^2}{\sum_{i=1}^n(Y_i-\mu_2)^2})$。【注：$\frac{1}{F_\alpha(n_2,n_1)}=F_{1-\alpha}(n_1,n_2)$，有些书上推导的时候也习惯于使用$\frac{\sum_{i=1}^n\left(\frac{Y_i-\mu_2}{\sigma_2}\right)^2/n_2}{\sum_{i=1}^n\left(\frac{X_i-\mu_1}{\sigma_1}\right)^2/n_1}\sim F(n_2,n_1)$，这样之后的$\frac{\sigma_1{}^2}{\sigma_2{}^2}$会呆在<中间<不动，方便得出结论】

(8).对于两个正态总体，待估参数为$\frac{\sigma_1{}^2}{\sigma_2{}^2}$，未知$\mu_1,\mu_2$：

利用定理 3 的(2).$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$，可得 $P\{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)<\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}<F_{\frac{\alpha}{2}}(n_1-1,n_2-1)\}=1-\alpha$，即 $P\{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}<F_{\frac{\alpha}{2}}(n_1-1,n_2-1)\frac{\sigma_1^2}{\sigma_2^2}\}=1-\alpha$，于是$\frac{\sigma_1{}^2}{\sigma_2{}^2}$的置信度为$1-\alpha$的双侧置信区间为
$(\frac{1}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)}\frac{S_1^2}{S_2^2},\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)}\frac{S_1^2}{S_2^2})$，即$(F_{1-\frac{\alpha}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2},F_{\frac{\alpha}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2})$。

(9).待估参数为 p，已知 n：

根据列维-林德博格定理，或者直接用棣莫弗-拉普拉斯定理：可得
$\lim_{n\to\infty}P\{\frac{\sum_{i=1}^n X_i-np}{\sqrt{np(1-p)}}<x\}=\Phi(x)$，$\frac{\sum_{i=1}^n X_i-np}{\sqrt{np(1-p)}}$分子分母同时除以 n 可得：$\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}}\dot{\sim}N(0,1)$，因此 $P\{z_{1-\frac{\alpha}{2}}<\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}}<z_{\frac{\alpha}{2}}\}=1-\alpha$。于是$(\overline{X}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}},\overline{X}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}})$，又因 $E(\overline{X})=p$，所以置信区间为$(\overline{X}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}},\overline{X}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}})$。其中$\overline{X}$可被写为$\frac{n_A}{n}$，即频

率。【由于只有这一点用了近似分布$\dot{\sim}$，所以用p的无偏估计量$\overline{X}$来近似p就不是不可原谅了：反正都已经是近似了】

(二).单侧置信区间

有时候人们只关心参数在一个方向的界限，对于设备元件的使用寿命，过长没问题，过短就有问题，此时我们关心其置信度为$1-\alpha$的单侧置信下限；而对于药品保质期或者毒性，我们就研究其单侧置信上限。此时取而代之 $P\{\hat{\theta}_1<\theta<\hat{\theta}_2\}=1-\alpha$的是，$P\{\theta<\hat{\theta}_2\}=1-\alpha$以及 $P\{\hat{\theta}_1<\theta\}=1-\alpha$。那么可见我们之前取的$t_{\frac{\alpha}{2}}$等，得变成$t_{\alpha}$了：

(1).假设我们要估计一个正态总体 X 的分布(密度)函数中的μ值，若已知σ^2：

则之前的 $P\{z_{1-\frac{\alpha}{2}}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<z_{\frac{\alpha}{2}}\}=1-\alpha$。会被修正为 $P\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<z_{\alpha}\}=1-\alpha$或者 $P\{-z_{\alpha}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\}=1-\alpha$，于是最终得到$\mu$的置信度为$1-\alpha$的单侧置信区间：$(\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_{\alpha},+\infty)$以及$(-\infty,\overline{X}+\frac{\sigma}{\sqrt{n}}\cdot z_{\alpha})$。

(2).同样对于单个正态总体，同样要估计其μ值，不过现在未知σ^2：

对它 $P\{t_{\frac{\alpha}{2}}(n-1)<\frac{\overline{X}-\mu}{S/\sqrt{n}}<t_{\frac{\alpha}{2}}(n-1)\}=1-\alpha$进行修正后，可得$(\overline{X}-\frac{S}{\sqrt{n}}\cdot t_{\alpha}(n-1),+\infty)$和$(-\infty,\overline{X}+\frac{S}{\sqrt{n}}\cdot t_{\alpha}(n-1))$。

(3).对于单个正态总体，待估参数为σ^2，现已知μ：

可得$(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi_{\alpha}^2(n)},+\infty)$、$(0,\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi_{1-\alpha}^2(n)})$。

这里要注意σ^2的取值范围，同样，$\frac{{\sigma_1}^2}{{\sigma_2}^2}$的取值范围也是$(0,+\infty)$。

第八章 假设检验

8.1 假设检验问题

分为参数假设检验(即总体分布已知、参数未知)，非参数假设检验(总体分布未知)。对总体分布中某些参数，或者总体的分布形式，提出假设，然后根据总体的样本对假设进行检验，并做出接受或是拒绝假设的选择，这就是假设检验。

8.1.2 假设检验的基本思想

根据总体的样本，来推断关于总体的一个假设是否成立的方法：对于给定的问题(比如问生产线是否工作正常——即对于理想的μ_0，是否有样本的(即总体的)$\mu=\mu_0$；或者希望新工艺下的材料强度均值越大越好，即是否有$\mu\geq\mu_0$)，先提出两个假设(命题)，一个为原假设H_0(比如$\mu=\mu_0$；$\mu\leq\mu_0$)，另一个为与该假设对立的备选假设H_1($\mu\neq\mu_0$；$\mu>\mu_0$)。实际工作中，往往把不会被轻易否定的命题作为原假设，比如以上例子中，虽然提问$\mu\geq\mu_0$，但原假设为$\mu\leq\mu_0$，即不认为旧工艺能被新工艺取代。

(1).以$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，这个为例，我们首先开发出来这个东西时，其中只有$\overline{X}$为变量(可以不认为$n$是)，即我们想以此来估计$\overline{X}$的概率分布；之后，$\overline{X}$像$\sigma$一样变成了已知，该轮到$\mu$变成未知数了，此时我们想以此分布来估计$\mu$的置信区间(依概率的波动范围)；而到了这里，我们先假设原假设H_0成立，即$\mu=\mu_0$，此时连μ也固定为μ_0了，$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$中不再有任何变量，此时得到的不是依概率分布的一个区间而是一个数，这个数已然出现并落在区间内，而若其出现位置不在原假设H_0以较大概率所预言的区间内(比如得到真实样本$\overline{X}$，而原假设的μ_0所对应的$|\overline{X}-\mu_0|$太大，以至于$|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|$超过了显著性水平$\alpha$所对应的$z_{\frac{\alpha}{2}}$)，则我们以较大的概率不相信原假设$H_0$所说的$\mu=\mu_0$；$|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|>z_{\frac{\alpha}{2}}$的此时，认为$\overline{X}$与$\mu_0$的差异是显著的，否则认为差异是不显著的。

——即用到了带有概率性质的反证法：一般的反证法为，原假设成立的条件下导出的结论是绝对成立的，此时若事实与结论矛盾，则完全否定原假设。——而这里却：原假设成立所导出的结论，是小概率不成立的，但若这种小概率发生了，则原假设是小概率可信的。这个小概率就称为显著性水平，用α表示。

在关于总体期望的假设检验中，样本均值$\overline{X}$是对总体期望μ的无偏估计。如果H_0即$\mu=\mu_0$成立，则样本均值$\overline{X}$该服从 $N(\mu,\frac{\sigma^2}{n})=N(\mu_0,\frac{\sigma^2}{n})$，其取值$\overline{x}$应该集中在$\mu_0$附近。所以如果$|\overline{X}-\mu_0|$过大，$H_0$的正确性就值得怀疑。因此$|\overline{X}-\mu_0|$的大小可以用来检验假设$H_0$是否成立。——当$|\overline{X}-\mu_0|<c$(c 与$z_{\frac{\alpha}{2}}$相关，而$z_{\frac{\alpha}{2}}$由$\alpha$决定，因此$\alpha$会影响选择)时，我们就选择接受$H_0$；当$|\overline{X}-\mu_0|\geq c$时，就拒绝$H_0$。————【这里我们选择相信样本均值，怀疑$H_0$，即认为样本均值以大概率反映真$\mu$；之后在提及两个错误的第一个错误时，我们选择相信$H_0$，而怀疑样本均值，即认为作为判据的仅仅是个样本，而样本有随机性。即使它大概率反映$\mu\neq\mu_0$时的真μ，也有小概率反映的是$\mu=\mu_0$时的错误μ】

而当H_0即$\mu=\mu_0$成立时，又有$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$，因此$|\overline{X}-\mu_0|\geq c$，等效于$|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|\geq k$。$p\{|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|\geq k\}=\alpha$设定为很小(可见其中的 $k=z_{\frac{\alpha}{2}}$)，即在H_0成立的条件下，$\{|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|\geq z_{\frac{\alpha}{2}}\}$为小概率事件，若所信任的(能大概率代表真$\mu$的)$\overline{X}$却使得，$H_0$成立的条件下的小概率事件$\{|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|\geq z_{\frac{\alpha}{2}}\}$发生了($|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}|\geq z_{\frac{\alpha}{2}}$所确定的关于$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$的区域，叫做$H_0$的拒绝域)，那么我们应怀疑$H_0$的正确性，并且因$|\overline{X}-\mu_0|\geq c$而认为$\overline{X}$与$\mu_0$的差异是显著的。——当显著性水平$\alpha$增大时，$z_{\frac{\alpha}{2}}$减小，$\overline{X}$更可能落在$H_0$所大概率预言的区间$|\overline{X}-\mu_0|<c$之外，$H_0$更容易被拒绝；当$\alpha$取得较小时，$H_0$的接受域随$z_{\frac{\alpha}{2}}$扩大，拒绝域缩小，$H_0$更容易被接受。

(2).对于之前给的H_1：$\mu\neq\mu_0$，称为双边备选假设。因而检验该组假设H_0：$\mu=\mu_0$、H_1：$\mu\neq\mu_0$的假设检验，称为双边检验。对应的还有检验假设H_0：$\mu\leq\mu_0$、H_1：$\mu>\mu_0$的假设检验，称为右边检验。右边检验和左边检验统称为单边检验。

仍以$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$为例，若$H_0$：$\mu\leq\mu_0$、$H_1$：$\mu>\mu_0$，则当假设$H_0$成立时，$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\leq\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$，即 $P\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\geq k\}\leq P\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\geq k\}=\alpha$，即有 $P\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\geq k\}\leq\alpha$，使之成立的最小 k 值为$z_\alpha$，于是$H_0$的拒绝域为$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\geq z_\alpha$。——同样可得到$H_0$：$\mu\geq\mu_0$的拒绝域：$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\geq\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$，即$P\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\leq k\}\leq P\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\leq k\}=\alpha$，即有 $P\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\leq k\}\leq\alpha$，即有 $1-P\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\geq k\}\leq\alpha$，即 $P\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\geq k\}\geq 1-\alpha$，于是使之成立的最大 k 值为$z_{1-\alpha}$，代入 $P\{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\leq k\}\leq\alpha$，则$H_0$的拒绝域为$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\leq z_{1-\alpha}$，即$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\leq -z_\alpha$。【这里或许可以用$\mu_0$作为主体来理解，$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\geq z_\alpha$对应的是$\mu_0\leq\cdots$，这与$H_0$：$\mu\leq\mu_0$中的$\mu_0\geq\cdots$相矛盾，因此是$H_0$的拒绝域；这里我们将真值$\mu$的下限看作了$\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_\alpha$】

因此对于所有的H_0：$\mu\leq\mu_0$下的 $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$($\sigma^2$已知)、$T=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$($\sigma^2$未知)，原假设的拒绝域均是反号的，即均是≥符号的：$Z\geq z_\alpha$，$T\geq t_\alpha(n-1)$。这是因为 Z 和 T 中的分子所含的μ_0和μ前均有负号。——同样，对于H_0：$\sigma^2\leq\sigma_0^2$下的$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$($\mu$未知)，原假设的拒绝域也是反号的：$\chi^2\geq\chi_\alpha^2(n-1)$，这是因为$\sigma_0^2$在分母。(关于$\sigma^2\leq\sigma_0^2$，没有$\mu$已知的

情形，可能这是因为之前在(3).中所用的$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$或者$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$没有$S^2$？）【这里所谓的反号，是指针对$\mu$及其对应的 Z、T 为主体】

另外，对于两个正态总体，对于H_0：$\mu_1-\mu_2\leq\delta$下的 Z=$\frac{\overline{X}-\overline{Y}-\delta}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}}$(${\sigma_1}^2$,${\sigma_2}^2$已知)和 T=$\frac{\overline{X}-\overline{Y}-\delta}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ (${\sigma_1}^2$=${\sigma_2}^2$=σ^2未知)，原假设的拒绝域也是反号的：Z≥z_α、T≥$t_\alpha(n_1+n_2-2)$。原因也是分子中的δ前有个负号；对于H_0：${\sigma_1}^2\leq{\sigma_2}^2$，即$\frac{{\sigma_1}^2}{{\sigma_2}^2}\leq1$ 下的 F=$\frac{S_1^2/S_2^2}{1}$(μ_1,μ_2均未知)，原假设的拒绝域也是反号的：F≥$F_\alpha(n_1-1,n_2-1)$，这也是因为$\frac{{\sigma_1}^2}{{\sigma_2}^2}$=1 在分母。(关于$\sigma^2\leq\sigma_0^2$，没有$\mu$已知的情形，可能这是因为之前在(8).中所用的$\frac{\sum_{i=1}^{n}\left(\frac{X_i-\mu_1}{\sigma_1}\right)^2/n_1}{\sum_{i=1}^{n}\left(\frac{Y_i-\mu_2}{\sigma_2}\right)^2/n_2}$没有$S^2$？）

(3).参数的假设检验与参数的区间估计的方法有相似之处。在对同一个参数作假设检验与区间估计时，作区间估计的样本函数也用于作假设检验，只是它须在假设成立(即参数取为某个常数，比如$\mu=\mu_0$)的条件下才可以作为检验统计量。此外，样本函数的置信区间对应于(不是等于)假设H_0的接受域。

之前我们利用小概率原理，过分地相信了落入H_0的拒绝域的$\overline{X}$，认为H_0下的小概率事件在单次试验中不应发生，因此否定了H_0。但如果我们选择相信H_0，即H_0本身是正确的，只是由于在H_0下，样本值就恰好以其小概率落入了拒绝域，而拒绝了H_0。这种“拒真”的错误称为第一类错误。——相反，若零假设H_0本身是错的，仅由于样本值落入了接受域而接受了H_0，这种“取伪”的错误称为第二类错误。

根据检验法则，犯第一类错误的概率就是显著性水平α，这说明显著性水平可以控制犯第一类错误的概率。但当样本容量一定时，减小犯某类错误的概率，往往会增大犯另一类错误的概率。若要同时减小犯两类错误的概率，则除非增大样本容量。一般只限定犯第一类错误的概率α，而不考虑犯第二类错误的概率，这样的假设检验称为显著性检验。【这样看来，若$\overline{X}$落在H_0的拒绝域，且H_0真是错误时，则反观整个过程，我们不应用H_0的μ_0来给出$\overline{X}$落在H_0的拒绝域的概率α所对应的拒绝域，此时$\overline{X}$便不一定落在其拒绝域内，或者落在其内的概率α会依真μ的未知而异。因为这是在H_0成立的条件下才有的结论。——但若不这样的话，未知μ会使得问题更无从下手。】