概率论与数理统计讲义

中国科学技术大学统计与金融系概率统计教研室

2008年4月

目 录

第一章　事件与概率

教学目的：

1) 掌握随机事件的概念和相关运算.

2) 了解概率的不同定义，掌握古典概型的基本计算.

3) 掌握条件概率的概念，熟练运用全概率公式和Bayes公式.

4) 掌握事件独立的概念和有关运算.

§1.1　概率论发展简史

概率论起源于17世纪, 现在公认是1654年Pascal与Fermat就赌博中的数学问题所展开的讨论, 在讨论中提出了一些基本概念, 最典型的例子是如何分赌本的问题. 两个赌徒相约赌若干局, 谁先赢s局就算谁赢. 由此提出期望的概念. 之后几个数学大家Huygens, Bernouli, J, De Moivre 等研究了这个问题, Bernouli 对频率与概率接近这一事实给予了理论上的阐述. 1812年Laplace 在《分析概率论》中最早叙述了概率论的几个基本定理, 给出了古典概率的明确定义. 1814年在《概率的哲学探讨》一书中, 记载了一个有趣的统计故事, 根据伦敦、彼得堡、柏林和全法国的统计资料, 得出几乎一致的男婴和女婴出生的比例为22:21, 即男婴比例为51.16%, 或男婴与女婴的比值为104.76:100, 可是统计1745-1784年整整40年巴黎男婴的出生率时, 得到的比例为25:24 (104.17:100), 调查研究后发现巴黎人有遗弃男婴的陋习. 1900年Hilbert 在第二届世界数学家大会上提出了23个有名的问题, 主体是对新世纪数学发展方向的探讨. 关于建立概率论的公理体系是他所提的第六个问题"借助公理来研究那些在其中数学起重要作用的物理科学; 首先是概率和力学". 随后Poincare, Borel等都对概率论公理体系的建立做出了努力, 1933年苏联的大数学家Kolmogorov(1903-1987)正式提出了概率论的公理体系. 概率论从此得到迅速的发展, 在此基础上, 数理统计也得到了迅速的发展.

§1.2　概率论的几个基本概念

§1.2.1　随机试验和随机事件

随机现象: 自然界中的客观现象, 当人们观测它时, 所得结果不能预先确定, 而仅仅是多种可能结果之一.

举例说明随机现象.

随机试验: 随机现象的实现和对它某个特征的观测.

随机试验中要求试验的结果至少2个，每次试验或观测得到其中的一个结果，在试验和观测之前不能预知是哪个结果发生。此外，要求在相同的条件下能重复试验。

如观测把硬币抛4次后正面向上的次数; 观测某地的温度变化; 某电话总机单位时间内转接的电话次数.

定义 1.2.1. 基本事件: 随机试验中的每个单一结果, 它犹如分子中的原子, 在化学反应中不能再分, 所以有"基本"两字.

如把硬币抛3次后有8种可能结果: 正正正、正正反、正反正、反正正、正反反、反正反、反反正、反反反. 这8种可能结果的每一个都是基本事件.

定义 1.2.2. 随机事件: 简称事件, 在随机试验中我们所关心的可能出现的各种结果, 它由一个或若干个基本事件组成.

随机事件常用大写英文字母A, B, C, D等表示. 如果用语言表达, 则要用花括号括起来.

定义 1.2.3. 样本空间: 随机试验中所有基本事件所构成的集合, 通常用Ω或S表示.

例 1.2.1. 掷一枚骰子, 观察出现的点数. 则 $\Omega = \{1, 2, 3, 4, 5, 6\}$.

例 1.2.2. 考察某一地区的年降雨量, 则 $\Omega = \{x | 0 \leq x < T\}$, 这里 T 表示某个常数, 表示降雨量不会超过 T.

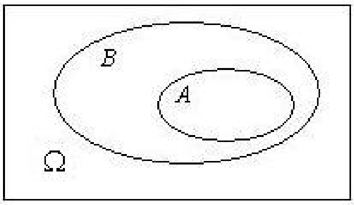
定义 1.2.4. 必然事件(Ω): 在试验中一定会发生的事件;

不可能事件(ϕ): 在试验中不可能发生的事件.

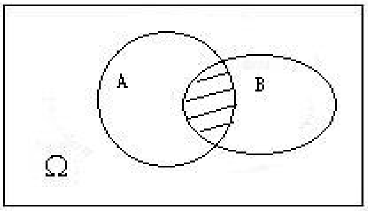
§1.2.2 事件的运算

可以证明, 把样本空间中的基本事件与空间中的点相对应, 则事件与集合相对应, 因此事件运算与集合运算可以建立一一对应关系.

1. 子事件$A \subset B$: 事件A发生蕴含事件B一定发生, 则事件A称为事件B的子事件, 记为$A \subset B$. 若$A \subset B$, 且$B \subset A$, 则称事件A与事件B相等, 记为$A = B$.

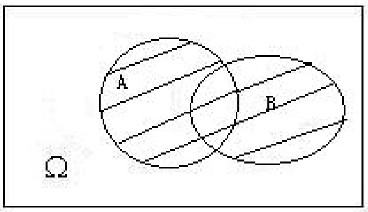


2. 事件的和($A \cup B$)：事件A和事件B中至少有一个发生的这一事件称为事件A和事件B的和, 记为$A \cup B$.

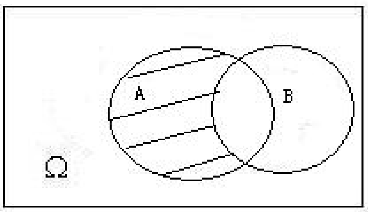


3. 事件的积($A \cap B$)：事件A和事件B同时发生这一事件称为事件A和事件B的积, 记为$A \cap B$.

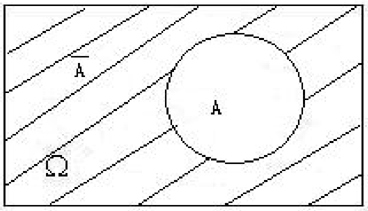
   如果$A \cap B = \phi$, 则称A和B不相容, 即事件A和B不能同时发生.



4. 对立事件A^c(或$\bar{A}$): A不发生这一事件称为事件A的对立事件(或余事件) .



5. 事件A和事件B的差$A-B$: 事件A发生而事件B不发生这一事件称为事件A和事件B的差, 记为$A-B$, 或等价的, AB^c.



De Morgan对偶法则:

$$\overline{A\cup B}=\bar{A}\cap\bar{B},$$

$$\overline{A\cap B}=\bar{A}\cup\bar{B},$$

上面公式可以推广到n个事件:

$$\overline{\bigcup_{i=1}^{n}A_i} = \bigcap_{i=1}^{n}\bar{A}_i$$
$$\overline{\bigcap_{i=1}^{n}A_i} = \bigcup_{i=1}^{n}\bar{A}_i$$

§1.2.3 概率的定义及性质

1. 概率的定义

什么叫概率? 直观地讲, 概率是随机事件发生可能性大小的数字表征, 其值在0和1之间, 换句话说, 概率是事件的函数. 如何求出事件A的概率(记为$P(A)$)?

(1) 古典概型: 有两个条件,

第一, (有限性) 试验结果只有有限个(记为n) ,

第二, (等可能性) 每个基本事件发生的可能性相同.

为计算事件A的概率, 设A中包含m个基本事件, 则定义事件A的概率为

$$P(A)=\frac{m}{n}$$

记号: 为方便起见，以$\#(B)$记事件B中基本事件的个数，因此，

$$P(A)=\frac{\#(A)}{\#(\Omega)}$$

(2) 概率的统计定义

古典概型的两个条件往往不能满足, 此时如何定义概率? 常用的一种方法是把含有事件A的随机试验独立重复做n次($Bernouli$试验) , 设事件A发生了n_A次, 称比值$\frac{n_A}{n}$为事件A发生的频率, 当n越来越大时, 频率会在某个值p附近波动, 且波动越来越小, 这个值p就定义为事件A的概率.

注意: 为什么不能写为$\lim_{n\to\infty}\frac{n_A}{n}=p$? 因为$\frac{n_A}{n}$不是$n$的函数.

几个例子: 英文字母被使用的频率是相当稳定的; 福尔摩斯探案集第四本《跳舞的小人》, 福尔摩斯用频率破了丘比特和埃尔茜之间联络密码; 1872年英国人$Shix,W$ 把π算到707位, 1944.5-1945.3数学家法格逊认为π的小数位的数字对0到9应该是等可能的, 但核对$Shix$的结果发现数字7太少, 故对$Shix$的结果有怀疑, 重新计算发现前527位是正确的, 后面不对了. 计算机出现后, 法国人让.盖尤计算了π的前100万位小数, 发现各个数字出现的频率相同.

(3) 主观概率

关于概率的统计定义, 我们可能会想到, 如果试验不能在相同的条件下独立重复很多次时该怎么办? 还有人们常谈论种种事件出现机会的大小, 如某人有80%的可能性办成某事. 如某人有80%的可能性办成某事.另一人则认为仅有50%的可能性. 即我们常常会拿一个数字去估计这类事件发生的可能性, 而心目中并不把它与频率挂钩. 这种概率称为主观概率, 这类概率有相当的生活基础. 在金融和管理等方面有大量的应用, 这一学派称为$Bayes$ 学派, 近来得到越来越多的认可. 但是当前用频率来定义概率的频率派仍是数理统计的主流. 焦点是频率派认为概率是客观存在，不可能因人而异.

(4) 概率的公理化定义

对概率运算规定一些简单的基本法则,

(i) 设A 是随机事件, 则$0\le P(A)\le 1$,

(ii) 设Ω为必然事件, 则$P(\Omega)=1$,

(iii) 若事件A和B不相容, 则$P(A\cup B)=P(A)+P(B)$,

为了对可数无穷个事件仍能成立, 我们要把上面公式中的两个事件推广到可数无穷个两两不相容的事件序列

$$P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$$

2. 古典概率计算的几个例子

计算古典概率, 主要用到排列组合的知识.

复习选排列, 重复排列和组合公式有关知识.

例 1.2.3. *一个班有r个人, 不计2月29日出生的(即假定一年为365天), 问至少有两人同一天生日的概率是多少?*

要点: (1) 本问题中的样本空间是什么? (2) 重复排列, (3) 先计算余事件

例 1.2.4. *盒中有32只红球, 4只白球, 从中任摸2球, 求两球中至少有一个白球的概率.*

要点: (1) 样本空间可以考虑为所有可能的组合, 也可以考虑为所有可能的选排列, 有些问题中只能考虑其中之一, 具体问题具体分析,

(2) 本题可以直接计算随机事件的概率, 也可以先计算对应的余事件的概率, 然后得到所需事件的概率.

§1.2.4 条件概率

1. 条件概率的定义

一般讲, 条件概率就是在知道了一定的信息下所得到的随机事件的概率. 如两个工厂A和B生产同一品牌的电视机, 商场中该品牌有个统一的次品率, 比如0.5%, 如果你从某个途径知道该商场的这批电视机是A厂生产的, 则你买到的电视机的次品率不再是0.5%, 而应该比0.5%要小, 这个概率就是条件概率, 即你在知道了这批电视机是A厂生产的附加条件下的概率就是条件概率.

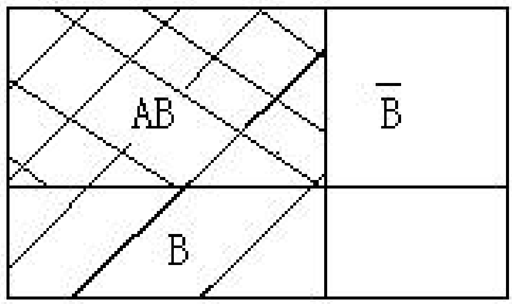
保险中应用的存活人数死亡率也是条件概率.

定义 1.2.5. *设事件A和B是随机试验Ω 中的两个事件, $P(B) > 0$, 称*

$$P(A|B) = \frac{P(AB)}{P(B)}$$

为事件B发生条件下事件A发生的条件概率.

注 1.2.1. *$P(A)$和$P(A|B)$ 是不同的两个概率. 如图, 设矩形A的面积为1, 则$P(A)$表示A的面积, 而$P(A|B)$表示在B中, A所占的比例, 即AB这块面积在B中所占的比例.*



也可以从概率的统计定义, 即用频率来近似概率这一角度来理解条件概率. 设在n次独立试验中, 事件A发生了n_A次, 事件B发生了n_B次, 事件AB发生了n_{AB}次, 事件B发生下事件A发生的频率为

$$\frac{n_{AB}}{n_B} \approx \frac{P(AB)}{P(B)}$$

注 1.2.2. *事实上, 我们所考虑的概率都是在一定条件下计算的, 因为随机试验就是在一定的条件下进行的, 所以样本空间是相对而言的. 如果把在一定条件下的事件试验看成无条件的, 则在补充条件下进行的事件试验的结果一般而言相对于原有结果要少, 即样本空间改变了. 所以所得随机事件的概率一般是不相同的.*

例 1.2.5. *有10个产品, 内有3个次品, 从中一个个地抽取(不放回) 检验, 问第一次取到次品后第二次再取到次品的概率.*

解: 样本空间Ω是从10个产品中有序取出2个产品的不同方法, 这是一个排列问题, 易知$\#\Omega = 10\times 9 = 90$, 记$A =${第一次取出的是次品}, $B =${第二次取出的是次品}, $\#(AB) = 6$, $\#A = 3$, 故

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{6/90}{3/10} = 2/9$$

注意, $P(B|A) = 2/9 \neq P(A) = 3/10$.

例 1.2.6. *有三张相同的卡片和一顶帽子, 第一张卡片两面都画有圈, 第二张卡片一面画圈, 一面画星, 第三张卡片两面都画星. 现在庄家把卡片放在帽中摇晃, 然后让你任取一张, 把它放在桌上, 设你看到卡片上面的图案为圈, 然后庄家与你打赌下面的图案与上面一样时算庄家赢, 不一样是为你赢. 请问这样的赌博是否是公平的?*

这是著名数学家, 信息论的创建者之一A. Weaver 设计的, 他曾在50年的《科学美国人》上介绍过这个例子. 请大家想一想, 很有意思.

例 1.2.7. *掷两个骰子, 观测出现的点数, 分别以x和y表示第一和第二颗骰子掷出的点数, 记$A=\{(x,y):x+y\geq 9\}$, $B=\{(x,y):x>y\}$, 求$P(A|B)$ 和$P(B|A)$.*

容易算出$P(A|B)=2/15$, $P(B|A)=1/3$, 这说明这两个条件概率不是一回事.

2. 乘法定理

由$P(A|B)=\frac{P(AB)}{P(B)}\Rightarrow P(AB)=P(A|B)P(B)$

由归纳法容易推广为n个事件同时发生的概率有如下公式:

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)\cdots P(A_n|A_1\cdots A_{n-1})$$

上面公式的右边看似麻烦, 其实在实际中很容易算出. 在没有给出n个事件之间相互关系时, 这是计算n个事件同时发生的一个重要公式.

例 1.2.8. *某人忘了某饭店电话号码的最后一个数字, 因而随意拨号, 问他三次之内拨通电话的概率.*

解: 令A_i={第i次打通电话}, $i=1,2,3$, 则

$$\begin{aligned} P(3\text{次内拨通电话}) &= P(A_1\cup A_2\cup A_3) \\ &= 1-P(\bar{A}_1\bar{A}_2\bar{A}_3) \\ &= 1-\frac{9}{10}\frac{8}{9}\frac{7}{8}=0.3 \end{aligned}$$

§1.2.5 全概率公式和Bayes公式

1. 全概率公式

定义 1.2.6. *设$B_1,B_2,\cdots B_n$ 是样本空间Ω 中的两两不相容的一组事件, 即$B_iB_j=\phi$, $i\neq j$, 且满足$\bigcup_{i=1}^n B_i=\Omega$, 则称$B_1,B_2,\cdots B_n$ 是样本空间Ω 的一个分割(又称为完备事件群,英文为partition).*

全概率公式:设$\{B_1, B_2, \cdots B_n\}$是样本空间Ω的一个分割, A为Ω中的一个事件, 则

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

目的: 有时不容易直接计算事件A的概率, 但是在每个B_i上A的条件概率容易求出.

注意: 应用中最重要的是验证$\{B_1, B_2, \cdots B_n\}$构成样本空间的一个分割.

例 1.2.9. *设某厂产品的一个零部件是由三家上游厂商供货的. 已知有一半是A 厂提供的, B厂商和C 分别提供25% . 已知厂商A 和B 的次品率都是2%, C 的次品率为4%, 从该厂产品中任取一个产品, 问该产品的这个零部件是次品的概率.*

解: 记B_i ={取到的产品是B_i 厂生产的}, $i = 1, 2, 3$, 易见B_1, B_2, B_3 构成样本空间的一个分割, 且$P(B_1) = 0.5$, $P(B_2) = P(B_3) = 0.25$, $P(A|B_1) = P(A|B_2) = 0.02$, $P(A|B_3) = 0.04$, 由全概率公式马上得到

$$P(A) = 0.02 \times 0.5 + 0.02 \times 0.25 + 0.04 \times 0.25 = 0.025$$

例 1.2.10. *一条家狗在野营后走失了, 猜想狗有三种可能去向:*

A: 它已回家,

B: 仍在原地啃骨头,

C: 已走失到附近的树林中去了.

从狗的习性可估计上述三种可能性分别为$1/4, 1/2, 1/4$. 一个小孩被派回去找狗, 如果狗仍在原地啃骨头, 小孩能找到的可能性为90%, 如果狗已走失到附近的树林中去了, 则小孩能找到的可能性为50% . 问小孩能找到狗的概率.

解: 可以分析得出狗的三种去向构成样本空间的一个分割, 小孩能在不同情况下找到狗的概率是条件概率, 如果狗已回家, 小孩能找到狗的概率为0. 由全概率公式可以算出小孩能找到狗的概率为$23/40 = 57.5\%$.

2. Bayes公式

设$\{B_1, B_2, \cdots B_n\}$是样本空间的一个分割, A为Ω中的一个事件, $P(B_i) > 0$, $i = 1, 2, \cdots, n$, $P(A) > 0$, 则

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

什么情况下用Bayes公式? 由公式知, 分母就是事件A的概率, 而分子和等式左边的条件概率中的条件正好反过来. 所以我们知道在因果关系互换时必须用Bayes公式.

例 1.2.11. *一种诊断某癌症的试剂, 经临床试验有如下记录: 有癌症病人阳性的概率为95%, 无癌症病人阴性的概率为95%. 现用这种试剂在某社区进行癌症普查, 设该社区癌症发病率为0.5%, 问某人反应为阳性时该人患癌症的概率.*

解: 设$A=$\{反应为阳性\}, $C=$\{被诊断者患癌症\}, 由题意,

$$P(A|C) = 0.95, P(\bar{A}|\bar{C}) = 0.95, P(C) = 0.005,$$

现在要算的是$P(C|A)$. 这是典型的因果关系互换, 只能用Bayes公式.

$$\begin{aligned} P(C|A) &= \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|\bar{C})P(\bar{C})} \\ &= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.05 \times 0.995} \\ &= 0.087 = 8.7\% \end{aligned}$$

这说明用该试剂进行普查, 准确性只有8.7%. 计算表明, 如果两次反应为阳性时患癌症的概率达到了64%.

§1.2.6 事件的独立性

为了计算两个事件同时发生的概率, 可以运用乘法定理, $P(AB) = P(A|B)P(B)$. 什么情况下$P(AB) = P(A)P(B)$? 即AB同时发生的概率等于两个事件单独发生概率的乘积? 为此我们有如下的定义:

定义 1.2.7. *设A, B是随机试验中的两个事件, 若满足$P(AB) = P(A)P(B)$, 则称事件A和B相互独立.*

关于独立的概念, 应该是从实际出发, 如果能够判断事件B的发生与否对事件A的发生与否不产生影响, 则事件A, B即为独立. 如把一个硬币掷两次, 观测正反面出现的

情况, $A=${第一次出现正面}, $B=${第二次出现正面}, $AB=${两次都出现正面}, 样本空间Ω有4个基本事件, $\#(AB)=1, \#(A)=2, \#(B)=2$, 故

$$P(AB)=1/4,\ P(A)P(B)=1/2\cdot 1/2=1/4$$

即事件A,B相互独立. 事实上, 我们容易判断第一次是否出现正面与第二次是否出现正面没有任何影响, 即独立的. 设$\tilde{A}$ 表示事件A发生和不发生之一, $\tilde{B}$ 表示事件B发生和不发生之一. 由独立性的定义可以推知$P(\tilde{A}\tilde{B})=P(\tilde{A})P(\tilde{B})$, (这儿一共4个等式) . 独立性的定义可以推广到n个事件.

定义 1.2.8. *设$A_1, A_2, \cdots A_n$是随机试验中的n个事件, 以$\tilde{A}_i$ 表示A_i或$\bar{A}_i$之一. 若满足*

$$P(\tilde{A}_1\tilde{A}_2\cdots\tilde{A}_n)=P(\tilde{A}_1)P(\tilde{A}_2)\cdots P(\tilde{A}_n),$$

则称事件列$A_1, A_2, \cdots A_n$ 相互独立.

(上面有2^n个等式)

注意: 上面等式等价于对$A_1, A_2, \cdots A_n$中的任意k个事件$A_{i_1}, A_{i_2}, \cdots, A_{i_k}$, $k=2,\cdots n$, 有

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$$

注意: 独立和不相容是不同的两个概念.

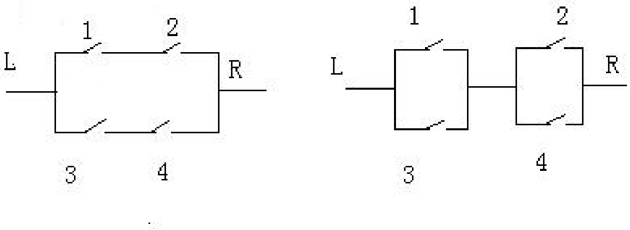
例 1.2.12. *A,B,C 三人独立地破译密码, 每人能破译密码的概率分别为$1/3, 1/4, 1/5$. 问密码能被破译的概率有多大?*

解: 设$D=${密码被破译}, A,B和C 分别表示A,B和C三人能破译密码这三个事件, 由独立性,

$$\begin{aligned} P(D) &= P(A\cup B\cup C)=1-P(\bar{A}\bar{B}\bar{C}) \\ &= 1-P(\bar{A})P(\bar{B})P(\bar{C})=1-\frac{2}{3}\frac{3}{4}\frac{4}{5}=0.6 \end{aligned}$$

例 1.2.13. *在元件可靠性研究中, 我们考虑如下两种电路:*

其中1-4表示4个继电器, 它们是否开通是相互独立的, 设继电器导通的概率为$p, (0<p<1)$, 求两种电路从L到R为通路的概率.



解: 左图为串联后并联, 右边为并联后串联, 记A_i ={第i个继电器导通}, 则左图LR为通路的表达为$A_1A_2 \cup A_3A_4$,右图LR为通路的表达为$(A_1 \cup A_3) \cap (A_2 \cup A_4)$, 由于$P(A_1A_2) = P(A_1)P(A_2) = p^2 = P(A_3A_4)$, 故

$$P(A_1A_2 \cup A_3A_4) = p^2 + p^2 - p^4 = p^2(2 - p^2)$$

同理,

$$P((A_1 \cup A_3) \cap (A_2 \cup A_4)) = (2p - p^2)^2 = p^2(2 - p)^2,$$

由于$2 - p^2 < (2 - p)^2$, 故并联后串联的电路比串联后并联的电路的可靠性高一点.

例 1.2.14. *n个人独立向同一目标射击, 第i个人命中目标的概率为$p_i, i = 1, 2, \cdots, n$, 求至少有一人命中目标的概率.*

解: 令A_i ={第i个人命中目标}, D ={至少有一人命中目标}, 则

$$\begin{aligned} D &= \bigcup_{i=1}^{n} A_i, \\ P(D) &= 1 - P(\bar{D}) = 1 - P(\bar{A}_1\bar{A}_2 \cdots \bar{A}_n) \\ &= 1 - (1 - p_1)(1 - p_2) \cdots (1 - p_n) \\ &\approx 1 - \exp\{-\Sigma p_i\} \end{aligned}$$

上面约等号在p_i较小时成立. 例如$p_i = 0.04$, $n = 100$ 时, $P(D) \approx 1 - \exp\{-4\} = 0.98168$

第二章　随机变量及其分布

教学目的:

1) 掌握随机变量的概念。掌握离散型随机变量的概率函数, 连续型随机变量的概率密度, 及任意的随机变量的分布函数的概念.

2) 掌握二项分布、Poisson分布，以及相应的概率计算.

3) 掌握正态分布，指数分布和均匀分布，会进行相应的概率计算.

4) 掌握多维随机变量的概念。了解n维随机变量的联合分布函数的概念和性质.

5) 掌握二维离散型和连续型随机变量的边缘分布与联合分布之间的关系，会用这些关系式求边缘分布.

§2.1　随机变量的概念

随机变量是其值随机会而定的变量。

例 2.1.1. *以X表示掷一次骰子得到的点数, X是一个随机变量. 它可以取$\{1,2,3,4,5,6\}$中的一个值, 但到底取那个值，要等掷了骰子才知道.*

例 2.1.2. *一张奖券的中奖金额是一个随机变量. 它的值要等开奖以后才知道.*

例 2.1.3. *在一批产品中随机地抽出100个产品, 其中所含的废品数是一个随机变量. 它的值要等检查了所有抽出的产品后才知道.*

在另外的例子中, 随机试验的结果虽然不是一个数, 但仍可用数来描述.

例 2.1.4. *掷一枚硬币出现正面或反面.*

例 2.1.5. *产品被分为正品或废品.*

上面两例中的结果均可用一个取值0,1的随机变量来描述, 其中可以1代表正面或正品, 以0代表反面或废品.

事实上, 对任意一个事件A, 定义

$$I_A(\omega) = \begin{cases} 1 & \omega \in A\ , \\ 0 & \text{反之}\ , \end{cases}$$

则事件A由随机变量I_A表示出来. I_A称为事件A的示性函数.

随机变量是把随机试验的结果，也就是样本空间，与一组实数联系起来. 这样的处理简化了原来的概率结构. 例如某机构调查民众对一提案的态度是支持(1)还是反对(0). 如果随机访问50人，按照古典概型，所有可能的结果有2^{50}个. 但是如果我们用X记1的个数来表示赞成者的人数，则X为一个随机变量. 它的取值范围只在$\{0, 1, \cdots, 50\}$. 所以随机变量的引进有利于我们对所研究的问题进行准确, 简练的描述. 又由于随机变量取实值, 随机变量之间的运算就变得容易了.

对于随机变量的研究，是概率论的中心内容. 因为对于一个随机试验，我们关心的通常是与所研究的问题有关的某个量或某些量. 而这些量就是随机变量.

定义 2.1.1. *令Ω为一个样本空间. 令X是定义在Ω上的一个实函数, 则称X为一个(一维)随机变量.*

常见的随机变量可以分为两大类. 只取有限个或可数个值的随机变量称为离散型随机变量；取连续的值且密度存在的随机变量称为连续型随机变量. 当然, 存在既非离散型也非连续型的随机变量. 但它们在实际中并不常见, 也不是我们这里研究的对象.

§2.2 离散型随机变量

定义 2.2.1. *设X为一随机变量. 如果X只取有限个或可数个值，则称X为一个(一维)离散型随机变量.*

由于一个随机变量的值是由试验结果决定的，因而是以一定的概率取值. 这个概率分布称为离散型随机变量的概率函数.

定义 2.2.2. *设X为一离散型随机变量，其全部可能值为$\{a_1, a_2, ...\}$. 则*

$$p_i = P(X = a_i),\ i = 1, 2, ... \tag{2.2.1}$$

称为X的概率函数.

概率函数$\{p_i,\ i=1,2,..\}$必须满足下列条件：

$$p_i \geq 0,\ i=1,2,....$$
$$\sum_i p_i = 1.$$

概率函数(2.2.1) 指出了全部概率1是如何在X的所有可能值之间分配的. 它可以列表的形式给出：

可能值	a_1	a_2	...	a_i	...
概率	p_1	p_2	...	p_i	...

(2.2.2)

有时也把(2.2.2)称为随机变量X的分布表.

设Ω为一样本空间. X为定义于其上的一个离散型随机变量，其取值为$x_1,x_2,....$. 令A为$\{x_1,x_2,...\}$的任意一个子集. 事件$\{X$取值于A中$\}$的概率可根据概率的可加性来计算：

$$P(A)=\sum_{x\in A}P(X=x).$$

这样知道了离散型随机变量X的概率函数，我们就能给出关于X的任何概率问题的回答.

下面我们给出常见的离散型分布. 在描述离散概率模型时, Bernoulli试验是最早被研究且应用及其广泛的概率模型.

定义 2.2.3. *设一个随机试验只有两个可能结果A和$\bar{A}$, 则称此试验为一Bernoulli试验.*

定义 2.2.4. *设将一个可能结果为A和$\bar{A}$的Bernoulli试验独立地重复n次, 使得事件A每次出现的概率相同, 则称此试验为n重Bernoulli试验.*

下面的0-1分布和二项分布都是以Bernoulli试验为基础的.

§2.2.1 0-1分布

设随机变量X只取0,1两值，$P(X=1)=p$，$P(X=0)=1-p$，则称X服从0-1分布或Bernoulli分布. 0-1分布是很多古典概率模型的基础.

§2.2.2　二项分布

设某事件A在一次试验中发生的概率为p. 现把试验独立地重复n次. 以X记A在这n次试验中发生的次数，则X取值$0, 1, ..., n$，且有

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n. \tag{2.2.3}$$

称X服从二项分布，记为$X \sim B(n, p)$.

从

$$\sum_{i=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1,$$

我们知道(2.2.3) 确实是一个概率函数.

为了考察这个分布是如何产生的，考虑事件$\{X = i\}$. 要使这个事件发生，必须在这n次试验的原始记录

$$AA\bar{A}A...\bar{A}A\bar{A}$$

中，有i个A, $n - i$个$\bar{A}$, 每个A有概率p而每个$\bar{A}$ 有概率$1 - p$. 又由于每次试验独立，所以每次出现A与否与其它次试验的结果独立. 因此由概率乘法定理得出每个这样的原始结果序列发生的概率为$p^i(1-p)^{n-i}$. 但是i个A和$n - i$个$\bar{A}$的排列总数是$\binom{n}{k}$，所以有i个A的概率是:

$$\binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \cdots, n.$$

一个变量服从二项分布有两个条件：一是各次试验的条件是稳定的，这保证了事件A的概率p在各次试验中保持不变；二是各次试验的独立性. 现实生活中有许多现象不同程度地满足这些条件. 例如工厂每天生产的产品. 假设每日生产n个产品. 若原材料质量，机器设备，工人操作水平等在一段时间内保持稳定，且每件产品是否合格与其它产品合格与否并无显著性关联，则每日的废品数服从二项分布.

§2.2.3　Poisson分布

设随机变量X的概率分布为

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \cdots, \ \lambda > 0, \tag{2.2.4}$$

则称X服从参数为λ的Poisson分布，并记$X \sim P(\lambda)$.

由于 e^λ 有级数展开式

$$e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + ... + \frac{\lambda^k}{k!} + ...$$

所以

$$\sum_{k=0}^{\infty} P(X=k)=1.$$

穆德和格雷比尔著的《统计学导论》给出了Poisson分布的如下推导.

假定体积为V的液体包含有一个大数目N的微生物. 再假定微生物没有群居的本能, 它们能够在液体的任何部分出现，且在体积相等的部分出现的机会相同. 现在我们取体积为D的微量液体在显微镜下观察，问在这微量液体中将发现x个微生物的概率是什么？我们假定V远远大于D. 由于假定了这些微生物是以一致的概率在液体中到处散布，因此任何一个微生物在D中出现的概率都是D/V. 再由于假定了微生物没有群居的本能，所以一个微生物在D中的出现，不会影响另一个微生物在D中的出现与否. 因此微生物中有x个在D中出现的概率就是

$$\binom{N}{x}\left(\frac{D}{V}\right)^{x}\left(1-\frac{D}{V}\right)^{N-x}. \tag{2.2.5}$$

在这里我们还假定微生物是如此之小, 拥挤的问题可以忽略不考虑, 即N个微生物所占据的部分对于体积D来说是微不足道.

在(2.2.5)中令V和N趋向于无穷, 且微生物的密度$N/V=d$保持常数. 将(2.2.5)式改写成如下形式：

$$\begin{aligned}&\frac{N(N-1)(N-2)...(N-x+1)}{x!N^{x}}\left(\frac{ND}{V}\right)^{x}\left(1-\frac{ND}{NV}\right)^{N-x}\\ =\ &\frac{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}...\right)\left(1-\frac{x-1}{N}\right)(Dd)^{x}\left(1-\frac{Dd}{N}\right)^{N-x}}{x!}.\end{aligned}$$

当N变成无限时其极限为

$$e^{-Dd}(Dd)^{x}/x! \tag{2.2.6}$$

令$Dd=\lambda$，则(2.2.6)和(2.2.4)的形式相同. 这一推导过程还证明了λ是x的平均数，因为所考察的一部分体积D乘以整个的密度d就给出了在D中所预计的平均数目.

当N很大，p很小且Np趋于一个极限时，Poisson分布是二项分布的一个很好的近似. 而在N未知时，Poisson分布更显得有用. 我们有下面的定理.

定理 2.2.1. *在n重Bernoulli试验中，以p_n代表事件A在试验中出现的概率，它与试验总数n有关. 如果$np_n\to\lambda$, 则当$n\to\infty$时,*

$$\binom{n}{k}p_n^{k}(1-p_n)^{n-k}\to\frac{\lambda^{k}}{k!}e^{-\lambda}. \tag{2.2.7}$$

例 2.2.1. 现在需要100个符合规格的元件. 从市场上买的该元件有废品率0.01. 考虑到有废品存在, 我们准备买$100+a$个元件使得从中可以挑出100个符合规格的元件. 我们要求在这$100+a$个元件中至少有100个符合规格的元件的概率不小于0.95. 问a至少要多大?

解: 令

$$A=\{\text{在}100+a\text{个元件中至少有}100\text{个符合规格的元件}\}.$$

假定各元件是否合格是独立的. 以X记在 $100+a$ 个元件中的废品数. 则X服从 $n=100+a$ 和 $p=0.01$ 的二项分布, 且

$$P(A)=\sum_{i=1}^{a}\binom{100+a}{i}(0.01)^i(0.99)^{100+a-i}.$$

上式中的概率很难计算. 由于 $100+a$ 较大而 0.01 较小, 且 $(100+a)(0.01)=1+0.01a\approx 1$, 我们以$\lambda=1$的Poisson分布来近似上述概率. 因而

$$P(A)=\sum_{i=1}^{a}e^{-1}/i!.$$

当$a=0,1,2,3$时, 上式右边分别为 $0.368, 0.736, 0.920$ 和 0.981. 故取 $a=3$ 已够了.

§2.2.4 离散的均匀分布

设随机变量X取值$a_1,a_2,...,a_n$, 且有

$$P(X=a_k)=\frac{1}{n},\ k=1,...,n. \tag{2.2.8}$$

则称X服从离散的均匀分布.

可以看出, 离散的均匀分布正是古典概型的抽象.

§2.3 连续型随机变量

离散随机变量只取有限个或可数无限个值，而连续型随机变量取不可数个值. 这就决定了不能用描述离散型随机变量的办法来刻划连续型随机变量.

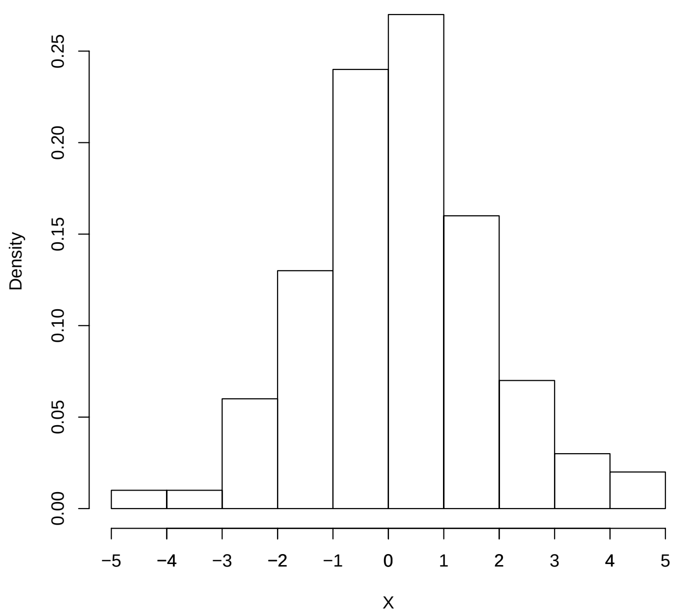
考虑一个例子. 假定步枪射手瞄准靶子在固定的位置进行一系列的射击. 令X是命中点与过靶心垂线的水平偏离值，设X取值$[-5cm,5cm]$. X是一个连续随机变量.

为了计算X落在某区间的概率，将$[-5,5]$分为长为1厘米的小区间. 对于每个小区间，以落在这个小区间的弹孔数除以弹孔总数得到落在这个区间的弹孔的相对频数. 设总弹孔数为100. 我们得到下表：

区间	弹孔数	相对频数
$[-5,-4]$	1	0.01
$[-4,-3]$	1	0.01
$[-3,-2]$	6	0.06
$[-2,-1]$	13	0.13
$[-1,0]$	24	0.24
$[0,1]$	27	0.27
$[1,2]$	16	0.16
$[2,3]$	7	0.07
$[3,4]$	3	0.03
$[4,5]$	2	0.02

上表可以用下图来表示：

图 2.3.1 弹孔位点分布图



我们注意每个矩形的底等于1，高为该矩形的区间所对应的相对频数，所以面积为相对频数. 全部矩形的面积是1. 对于$[-5,5]$的任一子区间，我们可以根据上图估计弹孔落在该子区间的概率. 例如要估计$0 < X \le 2$的概率，只要把区间中的两个矩形面积加起来，结果得到0.43. 再譬如说要估计$-0.25 < X \le 1.5$中的概率，我们应当计算该区间

上的面积，结果得到：

$$0.06 + 0.27 + 0.08 = 0.41.$$

如果第二批的100颗子弹射在靶子上，我们就将获得另一个经验分布. 它与第一个经验分布多半是不同的，尽管它们的外表可能相似. 如果把观察到的相对频数看作为某一“真”概率的估计，则我们假定有一个函数，它将给出任何区间中的精确概率. 这些概率由曲线下的面积给出. 由此我们得到如下定义：

定义 2.3.1. *X称为连续型随机变量，如果存在一个函数f，叫做X的概率密度函数，它满足下面的条件：*

1. *对所有的$-\infty < x < +\infty$, 有$f(x) \geq 0$;*
2. *$\int_{-\infty}^{+\infty} f(x)dx = 1$;*
3. *对于任意的$-\infty < a \leq b < +\infty$, 有$P(a \leq X \leq b) = \int_a^b f(x)dx$.*

注 2.3.1. *对于任意的$-\infty < x < +\infty$, 有$P(X = x) = \int_x^x f(u)du = 0$.*

注 2.3.2. *如果f只取某有限区间$[a, b]$的值, 令*

$$\tilde{f}(x) = \begin{cases} f(x) & x \in [a, b], \\ 0 & \text{其它}. \end{cases}$$

则$\tilde{f}$是定义在$(-\infty, +\infty)$上的密度函数, 且$f(x)$和$\tilde{f}(x)$给出相同的概率分布.

注 2.3.3. *假设有总共一个单位的质量连续地分布在$a \leq x \leq b$上. 那么$f(x)$表示在点x的质量密度且$\int_c^d f(x)dx$表示在区间$[c, d]$上的全部质量.*

由于连续随机变量的概率是用积分给出的, 我们可以直接处理密度的积分而不是密度本身.

定义 2.3.2. *设X为一连续型随机变量. 则*

$$F(x) = \int_{-\infty}^{x} f(u)du, \quad -\infty < x < +\infty \tag{2.3.1}$$

称为X的(累积)分布函数.

注 2.3.4. $F(x)$表示的是随机变量的数值小于或等于x的概率, 即

$$F(x) = P(X \leq x) \quad -\infty < x < +\infty. \tag{2.3.2}$$

由式(2.3.2)定义的F为X的(累积)分布函数的一般定义. 它适用于任意的随机变量. 设X为一离散型随机变量, 它以概率$\{p_1, ..., p_n, ..\}$取值$\{a_1, ..., a_n, ...\}$. 则

$$F(x) = \sum_{a_i \leq x} p_i.$$

分布函数F具有下列性质:

(1) F是非减的函数;

(2) $\lim_{x\to-\infty} F(x) = 0$;

(3) $\lim_{x\to+\infty} F(x) = 1$.

对于连续随机变量, 如果$F(x)$在点x的导数存在, 则

$$f(x) = F'(x).$$

连续随机变量的分布函数的图象如下图所示.

下面我们介绍常见的连续型分布. 它们包括正态分布, 指数分布和均匀分布.

§2.3.1 正态分布

如果一个随机变量X具有概率密度函数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < +\infty, \tag{2.3.3}$$

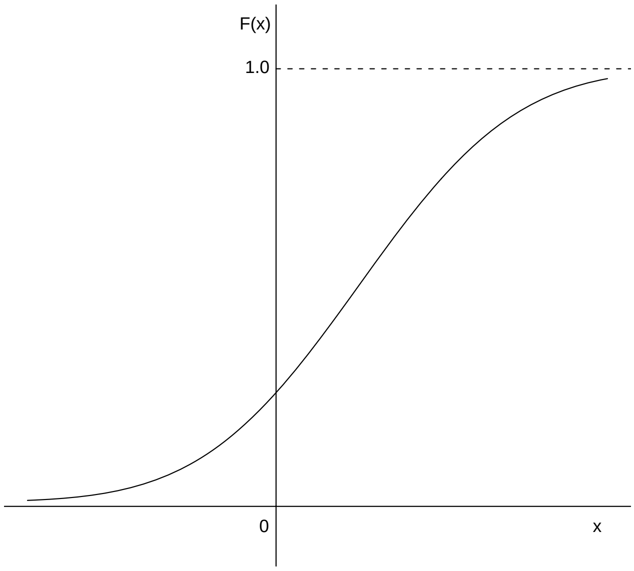
其中$-\infty < \mu < +\infty$, $\sigma^2 > 0$，则称X为一正态随机变量，记为$X \sim N(\mu, \sigma^2)$. 以(2.3.3)为密度的分布称为参数为μ和σ^2的正态分布.

具有参数$\mu = 0$, $\sigma = 1$的正态分布称为标准正态分布. 用$\Phi(x)$和$\phi(x)$表示标准正态分布$N(0, 1)$的分布函数和密度函数.

从图(2.3.3)可以看出, 正态分布的密度函数是以$x = \mu$为对称轴的对称函数. μ称为位置参数. 密度函数在$x = \mu$处达到最大值，在$(-\infty, \mu)$和$(\mu, +\infty)$内严格单调. 同时我们看到, σ的大小决定了密度函数的陡峭程度. 通常称σ为正态分布的形状参数.

以$F(x)$记正态分布$N(\mu, \sigma^2)$的概率分布函数，则恒有$F(x) = \Phi(\frac{x-\mu}{\sigma})$. 所以任一正态分布的概率分布函数都可通过标准正态分布的分布函数计算出来.

图 2.3.2 (累积)分布函数



例 2.3.1. 求数k使得对于正态分布的变量有$P(\mu - k\sigma < x < \mu + k\sigma) = 0.95$.

解: 令F为正态分布$N(\mu, \sigma^2)$的分布函数, 则有

$$P(\mu - k\sigma < x < \mu + k\sigma) = F(\mu + k\sigma) - F(\mu - k\sigma) = \Phi(k) - \Phi(-k) = 0.95. \tag{2.3.4}$$

从关系式 $\Phi(-k) = 1 - \Phi(k)$, 我们得 $2\Phi(k) - 1 = 0.95$. 所以$\Phi(k) = 0.975$. 查正态分布表, 得$k = 1.96$.

§2.3.2 指数分布

若随机变量X具有概率密度函数

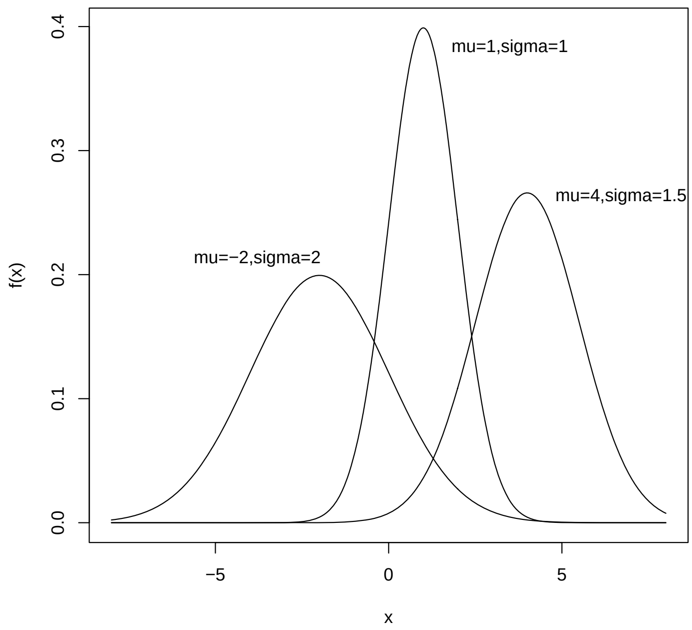
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x \leq 0, \end{cases} \tag{2.3.5}$$

其中$\lambda > 0$为常数, 则称X服从参数为λ的指数分布.

指数分布的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0, \\ 0 & x \leq 0. \end{cases} \tag{2.3.6}$$

图 2.3.3 正态分布的密度函数



从图(2.3.5)可以看出, 参数λ愈大, 密度函数下降得愈快.

指数分布经常用于作为各种"寿命"的分布的近似. 令X表示某元件的寿命. 我们引进X的失效率函数如下:

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x \le X \le x + \Delta x | X > x)}{\Delta x}.$$

失效率表示了元件在时刻x尚能正常工作, 在时刻x以后, 单位时间内发生失效的概率. 则如果

$$h(x) \equiv \lambda \quad (\text{常数}), \quad 0 < x < +\infty,$$

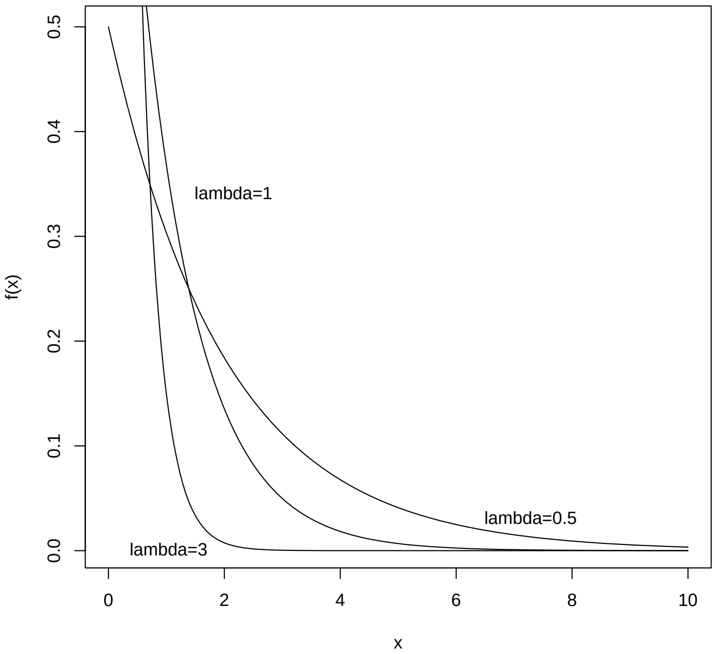
X服从指数分布. 即指数分布描述了无老化时的寿命分布.

指数分布的最重要的特点是"无记忆性". 即若X服从指数分布，则对任意的$s, t > 0$有

$$P(X > s + t \mid X > s) = P(X > t). \tag{2.3.7}$$

即寿命是无老化的. 可以证明, 指数分布是唯一具有性质(2.3.7)的连续型分布.

图 2.3.4 指数分布的密度函数



§2.3.3 均匀分布

设$a<b$，如果分布$F(x)$具有密度函数

$$f(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\ ,\\ 0 & \text{其它},\end{cases} \tag{2.3.8}$$

则称该分布为区间$[a,b]$上的均匀分布,记作$U[a,b]$. 如此定义的$f(x)$显然是一个概率密度函数. 容易算出其相应的分布函数为

$$F(x)=\begin{cases}0, & x\le a,\\ \frac{x-a}{b-a}, & a<x\le b,\\ 1, & x>b.\end{cases}$$

在计算时因四舍五入而产生的误差可以用均匀分布来描述.

§2.4 多维分布

在实际应用中，经常需要对所考虑的问题用多个变量来描述. 我们把多个随机变量放在一起组成向量，称为多维随机变量或者随机向量.

例 2.4.1. 从一付扑克牌中抽牌时, 可以用纸牌的花色和数字来说明其特征.

例 2.4.2. 考虑一个打靶的试验. 在靶面上取定一个直角坐标系. 则命中的位置可由其坐标(X,Y)来刻划. X,Y都是随机变量.

定义 2.4.1. 设$X = (X_1, \ldots, X_n)$. 如果每个X_i都是一个随机变量, $i = 1, \cdots, n$, 则称X为n维随机变量或者随机向量.

我们可以按照对常用一维随机变量的分类把常用的随机向量分为离散型和连续型的.

定义 2.4.2. 如果每一个X_i都是一个离散型随机变量, $i = 1, ..., n$, 则称$X = (X_1, \ldots, X_n)$为一n维离散随机变量. 设X_i的所有可能取值为$\{a_{i1}, a_{i2}, \cdots\}$, $i = 1, \ldots, n$, 则称

$$p(j_1, \cdots, j_n) = P(X_1 = a_{1j_1}, \ldots, X_n = a_{nj_n}),\ j_1, ..., j_n = 1, 2, ... \tag{2.4.1}$$

为n维随机变量X的概率函数.

容易证明概率函数具有下列性质:

(1) $p(j_1, \ldots, j_n) \geq 0, \quad j_i = 1, 2, \cdots,\ i = 1, 2, \ldots, n;$

(2) $\sum\limits_{j_1, \cdots, j_n} p(j_1, \ldots, j_n) = 1.$

我们具体来看一下二维离散分布. 设二维离散型随机变量(X,Y)的所有可能取值为$\{(x_i, y_j) : i = 1, ..., n, j = 1, 2, ..., m\}$. 我们经常以列联表的形式来表示二维离散型随机变量的概率分布. 记

$$p_{ij} = P(X = x_i, Y = y_j),\ i = 1, ..., n, j = 1, ..., m.$$

则(X,Y)的概率函数可以下表表示:

$Y \backslash X$	x_1	x_2	$\cdots$	x_n	行和
y_1	p_{11}	p_{21}	$\cdots$	p_{n1}	$p_{\cdot 1}$
y_2	p_{12}	p_{22}	$\cdots$	p_{n2}	$p_{\cdot 2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_m	p_{1m}	p_{2m}	$\vdots$	p_{nm}	$p_{\cdot m}$
列和	$p_{1\cdot}$	$p_{2\cdot}$	$\cdots$	$p_{n\cdot}$	1

例 2.4.3. 从一个包含五个黑球, 六个白球和七个红球的罐子里抽取四个球. 令X是抽到白球的数目, Y是抽到红球的数目. 则二维随机变量(X, Y)的概率函数为

$$p(x,y) = \frac{\binom{6}{x}\binom{7}{y}\binom{5}{4-x-y}}{\binom{18}{4}},\ 0 \le x+y \le 4. \tag{2.4.2}$$

以列联表表示, 即为

$Y \backslash X$	*0*	*1*	*2*	*3*	*4*	行和
0	$\frac{1}{612}$	$\frac{1}{51}$	$\frac{5}{102}$	$\frac{5}{153}$	$\frac{1}{204}$	$\frac{11}{102}$
1	$\frac{7}{306}$	$\frac{7}{51}$	$\frac{35}{204}$	$\frac{7}{153}$		$\frac{77}{204}$
2	$\frac{7}{102}$	$\frac{7}{34}$	$\frac{7}{68}$			$\frac{7}{17}$
3	$\frac{35}{612}$	$\frac{7}{102}$				$\frac{77}{612}$
4	$\frac{7}{612}$					$\frac{7}{612}$
列和	$\frac{99}{612}$	$\frac{22}{51}$	$\frac{11}{34}$	$\frac{4}{51}$	$\frac{1}{204}$	*1*

类似于一维连续型随机变量, 连续型随机向量的也是由密度函数来刻画的.

定义 2.4.3. 称$X = (X_1, \ldots, X_n)$为n维连续型随机变量, 如果存在$\mathbb{R}^n$上的非负函数$f(x_1, \ldots, x_n)$, 使得对任意的$-\infty < a_1 \le b_1 < +\infty, ..., -\infty < a_n \le b_n < +\infty$, 有

$$P(a_1 \le X_1 \le b_1, ..., a_n \le X_n \le b_n) = \int_{a_n}^{b_n} ... \int_{a_1}^{b_1} f(x_1, \ldots, x_n) dx_1 \cdots dx_n, \tag{2.4.3}$$

则称f为X的概率密度函数.

对n维随机变量我们也有分布函数的概念.

定义 2.4.4. 设$X=(X_1,\ldots,X_n)$为n维随机变量. 对任意的$(x_1,\ldots,x_n)\in\mathbb{R}^n$，称

$$F(x_1,\ldots,x_n)=P(X_1\le x_1,\ldots,X_n\le x_n) \tag{2.4.4}$$

为n维随机变量X的(联合)分布函数.

可以验证分布函数$F(x_1,\ldots,x_n)$具有下述性质:

(1) $F(x_1,\cdots,x_n)$对每个变元单调非降;

(2) 对任意的$1\le j\le n$有，$\lim\limits_{x_j\to-\infty}F(x_1,\cdots,x_n)=0$;

(3) $\lim\limits_{x_1\to\infty,\cdots,x_n\to\infty}F(x_1,\cdots,x_n)=1$.

对n维连续型随机变量, 从密度的定义我们有,

$$F(x_1,\ldots,x_n)=\int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1}f(x_1,\ldots,x_n)dx_1\ldots dx_n.$$

对高维离散型随机变量, 一般我们不使用分布函数.

例 2.4.4. 考虑二维随机变量$X=(X_1,X_2)$，其概率密度函数为

$$f(x_1,x_2)=\begin{cases}1/[(b-a)(d-c)] & \text{当 } a\le x_1\le b,\ c\le x_2\le d,\\ 0 & \text{其它}.\end{cases}$$

称此概率密度为$[a,b]\times[c,d]$上的均匀分布.

例 2.4.5. 设(X,Y)的概率密度函数有形式

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\}$$

其中$-\infty<a,b<\infty,\ 0<\sigma_1,\sigma_2<\infty,\ -1\le\rho\le1$. 称$(X,Y)$服从参数为$a,b,\sigma_1,\sigma_2,\rho$的二元正态分布，记为$N(a,b,\sigma_1^2,\sigma_2^2,\rho)$.

§2.5 边缘分布

设$(X_1,...,X_n)$为n维随机变量，其概率分布F已知. 令$X_{i_1},...,X_{i_m}$为$X_1,...,X_n$的任一子集，则$X_{i_1},...,X_{i_m}$的分布称为$X_1,...,X_n$或F的一个m维边缘分布.

我们先考虑离散型随机向量. 设二维离散随机变量(X,Y)的所有可能取值为$\{(x_i,y_j): i,j=1,2,\cdots\}$，则$(X,Y)$的联合分布律为

$$P(X=x_i,Y=y_j)=p_{ij} \quad i=1,...,n, j=1,2,...,m.$$

以列联表的形式表示就是

$Y \backslash X$	x_1	x_2	$\cdots$	x_n	行和
y_1	p_{11}	p_{21}	$\cdots$	p_{n1}	$p_{\cdot 1}$
y_2	p_{12}	p_{22}	$\cdots$	p_{n2}	$p_{\cdot 2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_m	p_{1m}	p_{2m}	$\vdots$	p_{nm}	$p_{\cdot m}$
列和	$p_{1\cdot}$	$p_{2\cdot}$	$\cdots$	$p_{n\cdot}$	1

从上述列联表我们可以计算随机变量X和Y的分布. 固定某个x_i. 因为Y在使得$X=x_i$的那些样本点上必取值为$y_1,...,y_m$中之一, 故有

$$p_X(x_i)=P(X=x_i)=\sum_j^m P(X=x_i,Y=y_j)=\sum_j^m p_{ij}=p_{i\cdot},\ i=1,2,\cdots n. \tag{2.5.1}$$

所以上述列联表的行和所表示的正是X的分布. 因为这个分布是从X和Y的联合分布推导出来的, 我们称(2.5.1)为X的边缘分布.

类似可以得到Y的边缘分布律

$$p_Y(y_j)=P(Y=y_j)=\sum_i^n p_{ij}=p_{\cdot j},\ j=1,2,\cdots m.$$

它是上述列联表的列和. 所以从列联表中, 我们不仅得到两个随机变量的联合分布, 同时通过将每行和每列相加, 得到两个变量的边缘分布.

类似地, 可对n $(n>2)$维的随机变量定义边缘分布. 设$X_1,...,X_n$为n维随机变量，其概率分布F已知. 令$X_{i_1},...,X_{i_m}$为$X_1,...,X_n$的任一子集，则$X_{i_1},...,X_{i_m}$的概率函数为

$$p_{i_1...i_m}(j_{i_1},...,j_{i_m})=P(X_{i_1}=a_{i_1 j_{i_1}},...,X_{i_m}=a_{i_m j_{i_m}})=\sum p(j_1,...,j_n).$$

其中和是对除$X_{i_1},...,X_{i_m}$之外的所有变量来求和.

现考虑连续型随机向量的边缘分布. 先考虑二维的情形. 设(X,Y)有概率密度函数$f(x,y)$. 则

$$\begin{aligned} P(x_1 \le X \le x_2) &= P(x_1 \le X \le x_2, -\infty < Y < +\infty) \\ &= \int_{-\infty}^{+\infty}\int_{x_1}^{x_2} f(u,v)dudv \\ &= \int_{x_1}^{x_2} f_X(u)du, \end{aligned} \tag{2.5.2}$$

其中

$$f_X(u) = \int_{-\infty}^{+\infty} f(u,v)dv. \tag{2.5.3}$$

从(2.5.2)我们可以看出, X的边缘密度函数即为(2.5.3). 类似地, Y的边缘密度函数为

$$f_Y(u) = \int_{-\infty}^{+\infty} f(u,v)du. \tag{2.5.4}$$

当$n > 2$时, 令$f(x_1,...,x_n)$为n维连续型随机变量$(X_1,...,X_n)$的概率密度函数. 设$(i_1, \cdots, i_m)$ 为$(1,2,...,n)$的一个子集. 则同上可证, $X_{i_1},...,X_{i_m}$的概率密度函数为

$$f(x_{i_1},...,x_{i_m}) = \int ... \int f(x_1,...,x_n)dx_1...dx_n.$$

其中积分是对除$X_{i_1},...,X_{i_m}$之外的所有变量来求积.

例 2.5.1. 设(X_1,X_2)服从$N(a,b,\sigma_1^2,\sigma_2^2,\rho)$.则可证明$X_1$的边缘分布为$N(a,\sigma_1^2)$，$X_2$的边缘分布为$N(b,\sigma_2^2)$.

例2.5.1说明了虽然n维随机变量$X=(X_1,...,X_n)$的分布可以唯一决定其所有的边缘分布，但边缘分布不足以决定X的联合分布.

§2.6 条件分布和随机变量的独立性

§2.6.1 条件分布

一个随机变量(或向量)的条件概率分布，就是在给定(或已知)某种条件(某种信息)下该随机变量(向量)的概率分布。

1. 离散型随机变量的条件分布

设(X,Y)为二维离散型随机变量，其全部的可能取值为$\{(x_i,y_j):i,j=1,2,\cdots\}$。记其联合分布律为

$$p_{ij}=P(X=x_i,Y=y_j),\qquad i,j=1,2,\cdots$$

若对给定的事件$\{Y=y_j\}$，其概率$P(Y=y_j)>0$，则称

$$P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots$$

为在给定$Y=y_j$的条件下X的条件分布律(概率函数)。类似的，若$P(X=x_i)>0$，则称

$$P(Y=y_j|X=x_i)=\frac{P(X=x_i,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_{i\cdot}},\quad j=1,2,\cdots$$

为在给定条件$X=x_i$下Y的条件分布律。

例 2.6.1. 设二维随机向量(X_1,X_2)的联合分布律如下所示:

X_1 \ X_2	-1	0	5	行和$p_{i\cdot}$
1	0.17	0.05	0.21	0.43
3	0.04	0.28	0.25	0.57
列和$p_{\cdot j}$	0.21	0.33	0.46	1.00

试求当$X_2=0$时，X_1的条件分布律。

解: 由联合分布律先算出两个边缘分布律$p_{i\cdot}$与$p_{\cdot j}$并填入表中，由此进一步算出条件分布律为:

$$P\{X_1=1|X_2=0\}=\frac{0.05}{0.33}=\frac{5}{33}$$

$$\text{而}\quad P\{X_1=3|X_2=0\}=\frac{0.28}{0.33}=\frac{28}{33}.$$

2. 连续型随机变量的条件分布

设(X,Y)有概率密度$f(x,y)$，我们考虑在给定$y\le Y\le y+\epsilon$的条件下X的条件分布函数(设$P\{y\le Y\le y+\epsilon\}>0$)

$$P(X\le x|y\le Y\le y+\epsilon)\ =\ \frac{P(X\le x,y\le Y\le y+\epsilon)}{P(y\le Y\le y+\epsilon)}$$

$$\begin{aligned} &= \int_{-\infty}^{x}\int_{y}^{y+\epsilon} f(u,v)dvdu \Big/ \int_{y}^{y+\epsilon} f_Y(y)dy \\ &= \int_{-\infty}^{x} \frac{\int_{y}^{y+\epsilon} f(u,v)dv}{\int_{y}^{y+\epsilon} f_Y(y)dy} du \end{aligned}$$

对上式两端关于x求导并令$\epsilon \to 0$, 可求得X在给定条件$Y=y$下的条件概率密度为

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad f_Y(y) > 0.$$

记为

$$X|y \sim f_{X|Y}(x|y).$$

类似地有Y在给定$X=x$的条件下的条件概率密度:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad f_X(x) > 0.$$

记为

$$Y|x \sim f_{Y|X}(y|x).$$

例 2.6.2. 设(X,Y)服从二元正态分布$N(a,b,\sigma_1^2,\sigma_2^2,\rho)$，试求$X|Y=y$的条件概率密度。

解:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\{-\frac{[x-(a+\rho\sigma_1\sigma_2^{-1}(y-b))]^2}{2\sigma_1^2(1-\rho^2)}\} \end{aligned}$$

即$X|Y=y \sim N(a+\rho\sigma_1\sigma_2^{-1}(y-b),\sigma_1^2(1-\rho^2))$。同理有：$Y|X=x \sim N(b+\rho\sigma_1^{-1}\sigma_2(x-a),\sigma_2^2(1-\rho^2))$。

例 2.6.3. 设X,Y服从单位圆上的均匀分布，试求$f_{X|Y}(x|y)$和$f_{Y|X}(y|x)$。

解: 由题设知(X,Y)的联合概率密度为

$$f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \le 1 \\ 0, & \text{其它} \end{cases}$$

易知

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & -1 \le x \le 1 \\ 0, & \text{其它} \end{cases}$$

所以

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{2\sqrt{1-y^2}}, & -\sqrt{1-y^2} \le x \le \sqrt{1-y^2} \\ 0, & \text{其它} \end{cases}$$

只需要把x, y互换，就可以得到$f_{Y|X}(y|x)$。

3. 更一般情形

无论离散型还是连续型条件分布，上述(X, Y)中的X和Y皆可推广到高维。例如: 设$(X_1, X_2, \cdots, X_n) \sim f(x_1, x_2, \ldots, x_n)$，且$(X_1, \cdots, X_k) \sim g(x_1, \ldots, x_k)$，则可定义在$(X_1, \cdots, X_k) = (x_1, \ldots, x_k)$的条件下，$(X_{k+1}, \cdots, X_n)$的条件密度为：

$$h(x_{k+1}, \ldots, x_n | x_1, \ldots, x_k) = \frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_k)}, \quad \text{其中 } g(x_1, \ldots, x_k) > 0.$$

注: 若记$(X_1, \cdots, X_k) = \boldsymbol{X}$，$(X_{k+1}, \cdots, X_n) = \boldsymbol{Y}$，$(x_1, \ldots, x_k) = \boldsymbol{x}$，$(x_{k+1}, \ldots, x_n) = \boldsymbol{y}$，则上式还可表示为:

$$h(\boldsymbol{y}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \boldsymbol{y})}{g(\boldsymbol{x})}, \ g(\boldsymbol{x}) > 0$$

§2.6.2 随机变量的独立性

若条件分布等于无条件分布，或者说条件分布与“条件”无关，例如，设$f_{X|Y}(x|y) = g(x)$，则可推出$g(x) = f_1(x)$，从而得到：

$$f(x, y) = f_1(x) f_2(y), \quad (x, y) \in \mathbb{R}^2$$

此时我们称X与Y是(相互)独立的。更一般的定义如下：

定义 2.6.1. 称离散型随机变量$X_1, \cdots, X_n$相互独立, 若它们的联合分布律等于各自的边缘分布律的乘积, 即

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n),$$

其中$(x_1, \ldots, x_n)$ 为$(X_1, X_2, \cdots, X_n)$ 的值域中的任意一点.

定义 2.6.2. 称连续型随机变量$X_1, \cdots, X_n$相互独立, 若它们的联合密度等于各自的边缘密度的乘积, 即

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n), \quad \forall \ (x_1, \ldots, x_n) \in \mathbb{R}^n$$

注: 更一般地, 有下面的的定义:

定义 2.6.3. 设$X_1, \cdots, X_n$为n个随机变量，如果它们的联合分布函数等于各自边缘分布函数的乘积，即

$$F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n), \quad \forall\ (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

则称随机变量$X_1, \cdots, X_n$相互独立.

在离散型和连续型两种情况下, 可以证明本定义分别与定义2.6.1和定义2.6.2等价.

例 2.6.4. 如果随机变量$X_1, \cdots, X_n$相互独立，则容易证明其中任何一部分随机变量也相互独立. 然而一般来说, 仅由某一部分独立却无法推出$X_1, \cdots, X_n$相互独立. 如见下例:

例 2.6.5. 若ξ, η相互独立,都服从-*1*和*1*这两点上的等可能分布，而$\zeta = \xi\eta$。则ζ, ξ, η两两独立但不相互独立。

例 2.6.6. 设$(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$，则$X$与$Y$相互独立的充要条件是$\rho = 0$。

例 2.6.7. 设(X, Y)服从矩形$D = [a, b] \times [c, d]$上的均匀分布，则X与Y相互独立。

例 2.6.8. 设(X, Y)服从单位圆上的均匀分布，则X与Y不独立。

例 2.6.9. 设有n个事件: $A_1, A_2, \cdots, A_n$，对于每个事件A_i，定义: $X_i = I_{A_i}$ (A_i的示性函数), $i = 1, 2, \cdots, n$，则可证明: $A_1, A_2, \cdots, A_n$ 独立$\Longleftrightarrow$ $X_1, X_2, \cdots, X_n$ 独立。

§2.7 随机变量的函数的概率分布

最简单的情形，是由一维随机变量X的概率分布去求其一给定函数$Y = g(X)$的分布。较常见的，是由$(X_1, X_2, \cdots, X_n)$的分布去求$Y = g(X_1, X_2, \cdots, X_n)$的分布。更一般地，由$(X_1, X_2, \cdots, X_n)$的分布去求$(Y_1, Y_2, \cdots, Y_m)$的分布，其中$Y_i = g_i(X_1, X_2, \cdots, X_n), \quad i = 1, 2, \cdots, m$。

这一部分内容，与数理统计中求统计量的分布有密切的联系。

1. 离散型随机变量的情形

设X的分布律为

$$P(X=x_i)=p_i, \quad i=1,2,\cdots$$

$g:R\to R$，令$Y=g(X)$，则Y的分布律为

$$P(Y=y_j)=P(g(X)=y_j)=\sum_{x_i:g(x_i)=y_j} P(X=x_i)=\sum_{i:g(x_i)=y_j} p_i$$

例 2.7.1. 设X的概率函数为

X	-1	0	1	2
P	1/4	1/2	1/8	1/8

试求$Y=X^2, \quad Z=X^3+1$的分布律。

解: 容易求得Y的分布律为:

Y	0	1	4
P	1/2	3/8	1/8

Z的分布律

Z	0	1	2	9
P	1/4	1/2	1/8	1/8

上述结论可以推广到多维随机变量的情形:

设随机向量X的分布律为$P(X=x)$，则X的函数$Y=g(X)$的分布律为

$$P(Y=y)=P(g(X)=y)=\sum_{x:g(x)=y} P(X=x)$$

特别当ξ,η是相互独立的非负整值随机变量，各有分布律$\{a_k\}$与$\{b_k\}$. 那么$\xi+\eta$有分布律

$$P(\xi+\eta=n)=\sum_{k=0}^{n} a_k b_{n-k}$$

称此公式为**离散卷积公式**

例 2.7.2. 设$X \sim B(n,p)$，$Y \sim B(m,p)$且X和Y相互独立，则$X+Y \sim B(n+m,p)$。这种性质称为再生性。可推广至多项和：设$X_i \sim B(n_i,p),(i=1,2,\cdots,m)$，且$X_1,X_2,\cdots,X_m$独立，则有:$\sum\limits_{i=1}^{m} X_i \sim B(\sum\limits_{i=1}^{m} n_i,p)$。特别，若$X_1,X_2,\cdots,X_n$为独立同分布，且$X_i \sim B(1,p)$, $i=1,\cdots,n$. 则有:$\sum\limits_{i=1}^{n} X_i \sim B(n,p)$。此结论揭示了二项分布与$0-1$分布之间的密切关系。

例 2.7.3. 设$X \sim P(\lambda)$，$Y \sim P(\mu)$，且X和Y独立，则有$X+Y \sim P(\lambda+\mu)$。即$Poisson$分布亦具有再生性，且可推广。

2. 连续型随机变量的情形

定理 2.7.1. *[密度变换公式]* 设随机变量X有概率密度函数 $f(x)$, $x \in (a,b)(a,b$可以为$\infty)$, 而$y=g(x)$在$x \in (a,b)$上是严格单调的连续函数，存在唯一的反函数$x=h(y), y \in (\alpha,\beta)$并且$h'(y)$存在且连续，那么$Y=g(X)$也是连续型随机变量且有概率密度函数

$$p(y)=f(h(y))|h'(y)|, \quad y \in (\alpha,\beta).$$

例 2.7.4. 设随机变量$X \sim U(-\frac{\pi}{2},\frac{\pi}{2})$, 求$Y=tgX$的概率密度函数。

由密度变换公式知Y的概率密度函数为

$$f(y)=\frac{1}{\pi}arctg'(y)=\frac{1}{\pi(1+y^2)}, \quad -\infty<y<\infty$$

此分布称为Cauchy分布。本题我们也可以用一般的方法求解，即先求出分布函数，然后对分布函数求导数得到。

$$\begin{aligned} F(y) &= P(Y \le y)=P(tg(X) \le y) \\ &= P(X \le arctg(y))=\int_{-\frac{\pi}{2}}^{arctg(y)} \frac{1}{\pi}dy=\frac{1}{\pi}arctg(y)+\frac{1}{2}. \end{aligned}$$

所以Y的概率密度为

$$f(y)=F'(y)=\frac{1}{\pi(1+y^2)}.$$

这种方法更具有一般性。

注: 当g不是在全区间上单调而是逐段单调时，密度变换公式为下面的形式:
设随机变量ξ的密度函数为$p_\xi(x)$, $a < x < b$. 如果可以把(a,b)分割为一些(有限个或可列个)互不重叠的子区间的和$(a,b)=\bigcup_j I_j$, 使得函数$u=g(t)$, $t\in(a,b)$在每个子区间上有唯一的反函数$h_j(u)$, 并且$h_j'(u)$存在连续, 则$\eta=g(\xi)$是连续型随机变量, 其密度函数为:

$$p_\eta(x)=\sum_j p_\xi(h_j(x))|h_j'(x)| \ . \tag{2.7.1}$$

例 2.7.5. 设$X\sim N(0,1)$，求$Y=X^2$的概率密度。

解: 由于函数$y=x^2$在$(-\infty,0)$和$[0,\infty)$上严格单调，因此由上述定理知Y的概率密度为

$$\begin{aligned} f(y) &= \phi(-\sqrt{y})|-\sqrt{y}'|I_{\{y>0\}}+\phi(\sqrt{y})|\sqrt{y}'|I_{\{y>0\}} \\ &= \frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}I_{\{y>0\}} \end{aligned}$$

定理 2.7.2. *设(ξ_1,ξ_2)是2维连续型随机向量, 具有联合密度函数$p(x_1,x_2)$, 设$\zeta_j=f_j(\xi_1,\xi_2), j=1,2$. 若$(\xi_1,\xi_2)$与$(\zeta_1,\zeta_2)$一一对应, 逆映射$\xi_j=h_j(\zeta_1,\zeta_2), j=1,2$.假定每个$h_j(y_1,y_2)$都有一阶连续偏导数. 则$(\zeta_1,\zeta_2)$亦为连续型随机向量, 且其联合概率密度为*

$$q(y_1,y_2)=\begin{cases} p\left(h_1(y_1,y_2),h_n(y_1,y_2)\right)|J|, & (y_1,y_2)\in\mathbb{D}, \\ 0, & (y_1,y_2)\notin\mathbb{D}, \end{cases} \tag{2.7.2}$$

其中$\mathbb{D}$是随机向量(ζ_1,ζ_2)的所有可能值的集合, J是变换的$Jaccobi$行列式, 即

$$J=\begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{vmatrix}$$

在多元随机变量场合，更一般地有

定理 2.7.3. *如果$(\xi_1,\cdots,\xi_n)$是n维连续型随机向量, 具有联合密度函数$p(x_1,\cdots,x_n)$. 假设存在n个n元函数*

$$y_j=f_j(x_1,\cdots,x_n),\quad j=1,\cdots,n,$$

使得

$$\zeta_j=f_j(\xi_1,\cdots,\xi_n),\quad j=1,\cdots,n,$$

若$(\xi_1,\cdots,\xi_n)$与$(\zeta_1,\cdots,\zeta_n)$之间一一对应，逆映射为$\xi_j=h_j(\zeta_1,\cdots,\zeta_n)$, $j=1,\cdots,n$. 其中每个$h_j(y_1,\cdots,y_n)$都有一阶连续偏导数，那么随机向量$(\zeta_1,\cdots,\zeta_n)$是连续型的，且具有联合密度函数

$$q(y_1,\cdots,y_n)=\begin{cases} p\left(h_1(y_1,\cdots,y_n),\cdots,h_n(y_1,\cdots,y_n)\right)|J|, & (y_1,\cdots,y_n)\in\mathbb{D}, \\ 0, & (y_1,\cdots,y_n)\notin\mathbb{D}, \end{cases} \tag{2.7.3}$$

其中$\mathbb{D}$是随机向量$(\zeta_1,\cdots,\zeta_n)$的所有可能值的集合，J是变换的$Jaccobi$行列式，即

$$J=\begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_n}{\partial y_n} \end{vmatrix}$$

例 2.7.6. 在直角坐标平面上随机选取一点，分别以随机变量ξ与η表示其横坐标和纵坐标，可以认为ξ与η相互独立. 如果ξ与η都服从正态分布$N(0,1)$，试求其极坐标(ρ,θ)的分布.

解: 易知

$$\begin{cases} x=r\cos t \\ y=r\sin t \end{cases}$$

是$(0,\infty)\times[0,2\pi)$与$\mathbb{R}^2$(原点除外)之间的一一变换, 变换的Jaccobi行列式

$$J=\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial t} \end{vmatrix}=\begin{vmatrix} \cos t & -r\sin t \\ \sin t & r\cos t \end{vmatrix}=r.$$

由于(ξ,η)的联合密度为

$$p(x,y)=\frac{1}{2\pi}\exp\left\{-\frac{x^2+y^2}{2}\right\},$$

所以由(2.7.3)式得知, (ρ,θ)的联合密度为

$$q(r,t)=\frac{1}{2\pi}r\exp\left\{-\frac{r^2}{2}\right\}=q_1(r)q_2(t),\quad r>0,\quad t\in[0,2\pi). \tag{2.7.4}$$

其中$q_1(r)=r\exp\left\{-\frac{r^2}{2}\right\},\quad r>0;\quad q_2(t)=\frac{1}{2\pi},t\in[0,2\pi)$.

这一结果表明: θ与ρ相互独立, 其中θ服从$[0,2\pi)$ 上的均匀分布; 而ρ则服从Weibull分布(参数$\lambda=1/2,\alpha=2$).

在计算两个随机变量之和时，我们还经常用到如下定理

定理 2.7.4. 设X,Y的联合概率密度为$f(x,y)$，则$X+Y$的概率密度$p(z)$为

$$p(z)=\int_{-\infty}^{\infty}f(x,z-x)dx=\int_{-\infty}^{\infty}f(z-y,y)dy$$

证一: 先求$X+Y$的分布函数$F(z)$. 我们有

$$\begin{aligned}F(z)=P(X+Y\le z)&=\int\int_{x+y\le z}f(x,y)dxdy=\int_{-\infty}^{\infty}dx\int_{-\infty}^{z-x}f(x,y)dy\\&=\int_{-\infty}^{\infty}du\int_{-\infty}^{z}f(x,t-x)dt=\int_{-\infty}^{z}\left\{\int_{-\infty}^{\infty}f(x,t-x)dx\right\}dt.\end{aligned}$$

这就说明,$X+Y$的分布函数$F(z)$是其中的花括弧中的函数在区间$(-\infty,z)$上的积分,所以$X+Y$是连续型随机变量,其密度函数如定理所述。

证二: 令$X=Z_1,X+Y=Z_2$, 利用单调映射的密度变换公式(2.7.2)可求得(Z_1,Z_2)的联合概率密度函数为$g(z_1,z_2)=f(z_1,z_2-z_1)$. 再对$g(z_1,z_2)$关于$z_1$在$R$上积分, 便求得$Z_2=X+Y$ 的密度为

$$\int_{-\infty}^{\infty}g(z_1,z_2)dz_1=\int_{-\infty}^{\infty}f(z_1,z_2-z_1)dz_1,$$

故得所证.

特别，当X与Y独立时，分别记X和Y的概率密度为$f_1(x)$和$f_2(y)$，则$X+Y$的概率密度为

$$p(z)=\int_{-\infty}^{\infty}f_1(x)f_2(z-x)dx=\int_{-\infty}^{\infty}f_1(z-y)f_2(y)dy\ \triangleq\ f_1*f_2(z)=f_2*f_1(z)$$

称此公式为**卷积公式**。

例 2.7.7. 设X服从期望为2的指数分布，$Y\sim U(0,1)$，且X和Y相互独立。求$X-Y$的概率密度和$P(X\le Y)$。

解一: 由题设知$-Y\sim U(-1,0)$，并记X和$-Y$的密度分别为f_1和f_2，从而由卷积公式有

$$\begin{aligned}f_{X-Y}(z)&=\int_{-\infty}^{\infty}f_1(x)f_2(z-x)dx\\&=\begin{cases}e^{-\frac{z}{2}}(1-e^{-\frac{1}{2}}), & z\ge 0\\ 1-e^{-\frac{z+1}{2}}, & -1<z<0\\ 0, & z\le -1\end{cases}\end{aligned}$$

所以$P(X\le Y)=P(X-Y\le 0)=2e^{-\frac{1}{2}}-1$。

解二: 由于

$$\begin{aligned} P(X-Y\le z) &= \int P(X\le z+y|Y=y)f(y)dy \\ &= \begin{cases} \int_0^1 P(X\le z+y)dy & z\ge 0 \\ \int_{-z}^1 P(X\le z+y)dy & -1<z<0 \\ 0 & z\le -1 \end{cases} \\ &= \begin{cases} 1-2e^{-z/2}(1-e^{-1/2}), & z\ge 0 \\ z+2e^{-(z+1)/2}-1, & -1<z<0 \\ 0, & z\le -1 \end{cases} \end{aligned}$$

再对分布函数求导数即得所求.

一些连续型随机变量，也有再生性性质。

例 2.7.8. 设$X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$且$X$与$Y$相互独立，则:

$$X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2).$$

更一般地, 设$X_i\sim N(\mu_i,\sigma_i^2), i=1,\cdots,n$, $X_1,\cdots,X_n$相互独立. $a_1,\cdots,a_n,b$为任意$n+1$个实数, 其中$a_1,\cdots,a_n$不全为零. 令$X=\sum\limits_{i=1}^n a_iX_i+b$, 则有: $X\sim N(\mu,\sigma^2)$, 其中$\mu=\sum\limits_{i=1}^n a_i\mu_i+b$, $\sigma^2=\sum\limits_{i=1}^n a_i^2\sigma_i^2$.

有时我们还会碰到计算随机变量之商的概率密度. 我们有

定理 2.7.5. 如果(ξ,η)是二维连续型随机向量, 它们的联合密度为$f(x,y)$, 则它们的商ξ/η是连续型随机变量, 具有密度函数

$$\begin{aligned} p_{\frac{\xi}{\eta}}(x) &= \int_{-\infty}^{\infty}|t|f(xt,t)dt, \quad \forall\ x\in\mathbb{R}. \\ \text{而}\ p_{\frac{\eta}{\xi}}(x) &= \int_{-\infty}^{\infty}|u|f(u,xu)du, \quad \forall\ x\in\mathbb{R}. \end{aligned} \tag{2.7.5}$$

例 2.7.9. 设随机变量ξ与η相互独立,同服从参数$\lambda=1$的指数分布, 试求ξ/η的密度函数.

解: 我们利用(2.7.5)式求$p_{\frac{\xi}{\eta}}(x)$.由于(ξ,η) 的联合密度为

$$p(u,v)=e^{-u-v},\quad u>0,\ v>0,$$

所以欲(2.7.5)式中的被积函数$|t|p(xt,t)\neq 0$, 当且仅当, $t>0$和$xt>0$, 从而知有

$$p_{\frac{\xi}{\eta}}(x)=\begin{cases}\int_0^{\infty} t\ e^{-xt-t}dt=\frac{1}{(1+x)^2} & x>0;\\ 0 & x\leq 0.\end{cases}$$

易见$p_{\frac{\eta}{\xi}}(x)$同上。

例 2.7.10. 设$X_1\sim N(0,1)$，$X_2\sim\chi_n^2$，且X_1与X_2独立。命$Y=\frac{X_1}{\sqrt{X_2/n}}$，则：$Y\sim t_n$(自由度为$n$的$t$分布)。

例 2.7.11. 设$X_1\sim\chi_m^2$, $X_2\sim\chi_n^2$, 且X_1与X_2独立, 则$Y=\frac{X_1/m}{X_2/n}\sim F_{m,n}$(自由度为$m,n$的$F$分布)。

例 2.7.12. 极小值和极大值的分布

对于n个随机变量$\xi_1,...,\xi_n$,我们可以考察它们的最大值和最小值:

$$\eta_1=\max\{\xi_1,...,\xi_n\},$$

$$\eta_2=\min\{\xi_1,...,\xi_n\}.$$

如此定义的η_1与η_2也是随机变量.

当$\xi_1,...,\xi_n$相互独立时, 我们不难利用它们的分布函数$F_1(x),...,F_n(x)$求出η_1与η_2 的分布函数$F_{\eta_1}(x)$和$F_{\eta_2}(x)$.

事实上,

$$\begin{aligned}F_{\eta_1}(x) &= P(\eta_1\leq x)=P(\max\{\xi_1,\cdots,\xi_n\}\leq x)=P\left(\bigcap_{k=1}^n(\xi_k\leq x)\right)\\ &= \prod_{k=1}^n P(\xi_k\leq x)=\prod_{k=1}^n F_k(x); \qquad (2.7.6)\end{aligned}$$

而利用关系式

$$(\eta_2>x)=(\xi_1>x,\ \dots,\xi_n>x)=\bigcap_{k=1}^n(\xi_k>x)$$

可得

$$\begin{aligned}F_{\eta_2}(x) &= P(\eta_2\leq x)=1-P(\eta_2>x)=1-P\left(\bigcap_{k=1}^n(\xi_k>x)\right)\\ &= 1-\prod_{k=1}^n P(\xi_k>x)=1-\prod_{k=1}^n(1-F_k(x)). \qquad (2.7.7)\end{aligned}$$

第三章　随机变量的数字特征

教学目的:

1) 理解随机变量的数学期望、方差的概念, 并会运用它们的基本性质计算具体分布的期望、方差.

2) 掌握二项分布、Poisson分布、均匀分布、指数分布、正态分布的数学期望和方差.

3) 会根据随机变量的概率分布计算其函数的数学期望.

4) 理解协方差、相关系数的概念, 掌握它们的性质, 并会利用这些性质进行计算, 了解矩的概念.

5) 理解大数定理与中心极限定理。

在前章中, 我们讨论了随机变量的概率分布, 这种分布是随机变量的概率论性质最完整的刻画. 而随机变量的数字特征是某些由随机变量的分布所决定的常数, 它刻画了随机变量或者说刻画了其分布的某一方面的性质, 这些性质往往是实际应用中人们比较关心的. 例如, 我们在了解某一行业工人的经济状况时, 我们首先关心的恐怕会是其平均收入, 这会给我们一个总体的印象, 而收入的分布状况, 倒不一定是最重要的, 这就是刻画总体平均值的数字特征. 另一类重要的数字特征, 是用来衡量随机变量取值的分散程度. 还拿我们上个例子说明, 如果我们考虑两个行业工人的经济状况, 他们的平均收入大体相近, 但是一个行业收入分配较平均, 即大多数人的收入都在平均值上下不远处, 分散程度就小; 另一个行业则相反, 其收入远离平均值很多, 分散程度就大, 这两者的实际意义当然很不相同. 平均值和分散度是刻画随机变量性质的两类最重要的数字特征. 除了这两者之外, 对于多维变量而言, 还有一类刻画各分量之间关系的数字特征, 较为常用的是协方差和相关系数, 这些我们将在下面的章节详细讨论. 数字特征另一个重要意义在于, 当我们不知道随机变量的确切概率分布, 但是清楚其数字特征的情形下, 我们可以根据这些数字特征推断该随机变量大致的概率性质. 比如某个工厂生产一批灯泡, 我们想了解这批灯泡的质量如何. 我们不知道这批灯泡寿命的确切概率分布, 但是如果我们知道这批灯泡的平均寿命, 知道这批灯泡寿命的分散程度, 那我们就可以大致推断出这批灯泡的质量状况.

§3.1 数学期望(均值)及中位数

§3.1.1 数学期望

数学期望也称均值, 是随机变量的一个最基本的数字特征. 我们先看如下的一个例子

例 3.1.1. *一甲乙两人赌技相同, 各出赌金100元, 约定先胜三局者为胜, 取得全部200元. 现在甲胜2局乙胜1局的情况下中止, 问赌本该如何分?*

解: 如果继续赌下去而不中止, 则甲有3/4的概率取胜, 而乙胜的概率为1/4. 所以, 在甲胜2局乙胜1局的这个情况下, 甲能期望"得到"的数目, 应当确定为

$$200 \times \frac{3}{4} + 0 \times \frac{1}{4} = 150(\text{元}),$$

而乙能"期望"得到的数目, 则为

$$200 \times \frac{1}{4} + 0 \times \frac{3}{4} = 50(\text{元}).$$

如果引进一个随机变量X, X等于在上述局面(甲值2胜乙1胜)之下, 继续赌下去甲的最终所得, 则X有两个可能的值: 200 和0, 其概率分别为3/4和1/4. 而甲的期望所得, 即X的"期望"值, 即等于

X的可能值与其概率之积的累加

这就是"数学期望"这个名称的由来. 另一个名称"均值" 形象易懂, 也很常用. 下面我们就给出数学期望(均值)的定义:

对一般的离散型分布, 我们有

定义 3.1.1. *设X为一离散型随机变量, 其分布律为*

$$P(X = x_i) = p_i, \quad i = 1, 2, \cdots$$

如果$\sum\limits_{i=1}^{\infty} |x_i| p_i < +\infty$, 则称

$$\sum_{i=1}^{\infty} x_i p_i$$

为随机变量X的数学期望(均值), 用符号EX表示. 若$\sum\limits_{i=1}^{\infty} |x_i| p_i = +\infty$, 则称$X$的数学期望(均值)不存在.

对连续型随机变量, 其数学期望的定义如下

定义 3.1.2. *如果连续型随机变量X具有密度函数$f(x)$, 则当*

$$\int_{-\infty}^{\infty}|x|f(x)dx<\infty$$

时, 我们将积分

$$\int_{-\infty}^{\infty}xf(x)dx$$

的值称为X的数学期望, 记作EX. 如果

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

则称X的数学期望不存在.

下面求解几种常见分布的数学期望.

1. 二项分布$X\sim B(n,p)$:

$$\begin{aligned}EX &= \sum_{k=0}^{n}k.\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\ &= np.\sum_{i=0}^{n-1}\frac{(n-1)!}{i!(n-1-i)!}p^i(1-p)^{n-1-i}\\ &= np\end{aligned}$$

2. Poisson 分布$X\sim P(\lambda)$:

$$EX=\lambda$$

3. 正态分布$X\sim N(\mu,\sigma^2)$:

$$\begin{aligned}EX &= \int_{-\infty}^{+\infty}\frac{x}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\ &= \int_{-\infty}^{+\infty}(\sigma y+\mu).\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy\\ &= \mu\end{aligned}$$

4. 均匀分布$X\sim U[a,b]$:

$$EX=\frac{a+b}{2}$$

5. 指数分布$X\sim Exp(\lambda)$:

$$EX=1/\lambda$$

§3.1.2 数学期望的性质

1. 若干个随机变量线性组合的期望, 等于各变量期望的线性组合. 假设$c_1, c_2, \ldots, c_n$为常数, 则有

$$E(c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_1EX_1 + c_2EX_2 + \cdots + c_nEX_n,$$

这里假定各变量的期望都存在.

例 3.1.2. 假设随机变量$X \sim B(n,p)$, 求EX.

解: 令$I_i \sim B(1,p)$, $i = 1, 2, \ldots, n$, 则$X = \sum_{i=1}^n I_i$ 且$EI_i = p$. 所以, $EX = \sum_{i=1}^n EI_i = np$.

2. 若干个独立随机变量之积的期望, 等于各变量的期望之积, 即

$$E(X_1X_2\cdots X_n) = EX_1EX_2\cdots EX_n,$$

这里假定各变量相互独立且期望都存在.

3. (随机变量函数的期望) 设随机变量X 为离散型, 有分布$P(X = a_i) = p_i$, $i = 1, 2, \ldots$, 或者为连续型, 有概率密度函数$f(x)$. 则

$$Eg(X) = \begin{cases} \sum_i g(a_i)p_i, & \sum_i |g(a_i)|p_i < \infty; \\ \int_{-\infty}^{+\infty} g(x)f(x)dx, & \int_{-\infty}^{+\infty} |g(x)|f(x)dx < \infty. \end{cases}$$

例 3.1.3. 假设c为常数, 则$EcX = cEX$.

例 3.1.4. 设随机变量$X \sim N(0,1)$, 求$Y = X^2 + 1$的数学期望.

解: 由$X \sim N(0,1)$, 我们有

$$\begin{aligned} EX^2 &= \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= 1. \end{aligned}$$

所以, $EY = EX^2 + 1 = 2$.

例 3.1.5. 飞机场载客汽车上有*20*位乘客, 离开机场后共有*10*个车站可以下车, 若某个车站没有人下车则该车站不停车. 设乘客在每个车站下车的可能性相等, 以X表示停车的次数, 求EX.

解: 设

$$Y_i = \begin{cases} 1, & \text{第 } i \text{ 个车站有人下车} \\ 0, & \text{第 } i \text{ 个车站无人下车} \end{cases} \qquad i = 1, \cdots, 20.$$

则显然$X = \sum\limits_{i=1}^{20} Y_i$, 所以

$$\begin{aligned} EX &= \sum_{i=1}^{20} EY_i = \sum_{i=1}^{20} P(\text{第 } i \text{ 个车站有人下车}) \\ &= \sum_{i=1}^{20} [1 - 0.9^{20}] = 8.784. \end{aligned}$$

§3.1.3 条件期望

我们知道条件分布也是一个概率分布, 因此类似数学期望的定义, 我们可以给出条件期望的定义. 在给定了随机变量X取值x的条件之下, Y的条件期望, 我们记为$E(Y|X=x)$, 也可简记为$E(Y|x)$.

定义 3.1.3. 设X和Y为随机变量, 若(X,Y)为离散型, 且在给定$X=x$之下, Y 有分布$P(Y=a_i|X=x)=p_i$, $i=1,2,\ldots$, 或者(X,Y)为连续型, 且在给定$X=x$之下, Y的条件密度函数为$f(y|x)$. 则

$$E(Y|X=x) = \begin{cases} \int_{-\infty}^{+\infty} y f(y|x) dy, & (X,Y) \text{ 为连续型;} \\ \sum_i a_i p_i, & (X,Y) \text{ 为离散型.} \end{cases}$$

期望所具有的性质条件期望同样满足.

例 3.1.6. 设$(X,Y) \sim N(a,b,\sigma_1^2,\sigma_2^2,\rho)$, 试计算$E(Y|X=x)$.

解: 由于$Y|X=x \sim N(b+\rho\frac{\sigma_2}{\sigma_1}(x-a), (1-\rho^2)\sigma_2^2)$, 所以由二维正态分布的性质知$E(Y|X=x) = b+\rho\frac{\sigma_2}{\sigma_1}(x-a)$.

[**注**]: 条件期望$E(Y|X=x)$是x的函数, 当我们将x换为X时, $E(Y|X)$ 就是一个随机变量.

我们有如下的公式成立:

定理 3.1.1. 设X,Y为两个随机变量. 则有

$$EX = E\{E[X|Y]\} \qquad [\text{全期望公式}]$$

证: 我们仅在连续型随机变量的情形下证明此定理. 设Y的p.d.f 为$p(y)$, $X|Y=y$的p.d.f为$q(x|y)$. 则

$$\begin{aligned} EX &= \iint_{-\infty}^{\infty} q(x|y)p(y)dxdy \\ &= \int\int_{-\infty}^{\infty} q(x|y)dxp(y)dy = \int_{-\infty}^{\infty} E[X|Y=y]p(y)dy \\ &= E\{E[X|Y]\} \end{aligned}$$

[**推广**]: 当$g(X)$ 为可积随机变量时, 有$Eg(X) = E\{E[g(X)|Y]\}$.

由此得到求解期望的第二种方法: 先求解$h(x) = E(Y|X = x)$, 再求解$Eh(X)$, 即可求得EY.

例 3.1.7. 一窃贼被关在有3个门的地牢里, 其中第一个门通向自由. 出这门走3个小时便可以回到地面; 第2个门通向另一个地道, 走5个小时将返回到地牢; 第3个门通向更长的地道, 走7个小时也回到地牢. 若窃贼每次选择3个门的可能性总相同, 求他为获得自由而奔走的平均时间.

解: 设这个窃贼需要走X小时才能到达地面, 并设Y代表他每次对3个门的选择情况, Y各以1/3的概率取值1, 2, 3. 则

$$EX = E[E(X|Y)] = \sum_{i=1}^{3} E(X|Y=i)P(Y=i)$$

注意到$E(X|Y=1) = 3, E(X|Y=2) = 5 + EX, E(X|Y=3) = 7 + EX$, 所以

$$EX = \frac{1}{3}[3 + 5 + EX + 7 + EX]$$

即得到$EX = 15$.

例 3.1.8. 设$(X,Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$, 试计算$EXY$.

解: 先算得

$$E(XY|X=x) = xE(Y|X=x) = x(b+\rho\frac{\sigma_2}{\sigma_1}(x-a));$$

所以

$$\begin{aligned} EXY &= E(bX+\rho\frac{\sigma_2}{\sigma_1}X^2-\rho\frac{\sigma_2}{\sigma_1}aX) \\ &= ab+\rho\frac{\sigma_2}{\sigma_1}(a^2+\sigma_1^2)-\rho\frac{\sigma_2}{\sigma_1}a^2 \\ &= ab+\rho\sigma_1\sigma_2. \end{aligned}$$

§3.1.4 中位数

我们已经知道, 随机变量X的数学期望就是它的平均值, 因此从一定意义上, 数学期望刻画了随机变量所取之值的"中心位置". 但是, 我们也可以用别的数字特征来刻画随机变量的"中心位置". 中位数就是这样一种数字特征.

定义 3.1.4. *称μ为连续型随机变量X的中位数, 如果*

$$P(X\le\mu)=\frac{1}{2}, \quad P(X\ge\mu)=\frac{1}{2}.$$

从定义上可以看出, m这个点把X 的分布从概率上一分两半: 在m左边占一半, m右边也占一半, 从概率上说, m 这个点正好居于中央, 这就是"中位数" 得名的由来. 在实用上, 中位数用得很多, 特别有不少社会统计资料, 常拿中位数来刻化某种量的代表性数值, 有时它比数学期望更说明问题, 例如, 某社区内人的收入的中位数告诉我们: 有一半人的收入低于此值, 另一半高于此值. 我们直观上感觉到这个值对该社区的收入情况, 的确很具有代表性, 和期望值相比它的一个优点是受个别特别大或特别小的值的影响很小, 而期望则不然, 举例而言, 若该社区中有一个收入在百万元以上, 则该社区的均值可能很高, 而绝大多数人并不富裕, 这个均值并不很有代表性, 中位数则不然, 它几乎不受少量这种特大值的影响.

从理论上说, 中位数与均值相比还与一个优点, 即它总存在, 而均值则不是对任何随机变量都存在. 虽则中位数有这些优点, 但在概率统计中, 无论理论和应用上, 数学期望的重要性都超过中位数, 其原因有一下两个方面:

1. 均值有很多优良的性质, 这些性质时使得在数学处理上很方便. 例如, $E(X_1+X_2)=EX_1+EX_2$, 而X_1+X_2 的中位数与X_1, X_2各自的中位数之间, 不存在简单的联系, 这使中位数在数学上的处理很复杂且不方便;

2. 中位数本身固有的某些缺点：中位数可以不唯一, 且对于离散型随机变量不易定义.

例 3.1.9. *设随机变量$X \sim B(1, \frac{1}{2})$, 求X的中位数.*

解: 由于X的分布函数为

$$F(x) = \begin{cases} 0, & x \le 0 \\ \frac{1}{2}, & 0 < x < 1 \\ 1, & x \ge 1 \end{cases}$$

由中位数的定义知区间(0,1)内的每一个数都是X的中位数, 所以此例说明中位数可以不唯一.

§3.2 方差、标准差和矩

§3.2.1 方差和标准差

现在我们转到本章开始时候提到的另一类数字特征, 即刻画随机变量在其中心位置附近散布程度的数字特征, 其中最重要的是方差. 在实际应用中, 方差不仅是信息度量的标准也是风险度量的标准.

定义 3.2.1. *设X 为随机变量, 分布为F, 则称*

$$Var(X) = E(X - EX)^2 = \sigma^2$$

为X (或分布F)的方差, 其平方根$\sqrt{Var(X)} = \sigma$ (取正值) 称为X (或分布F)的标准差.

显然有

$$Var(X) = EX^2 - (EX)^2.$$

对随机变量的方差, 我们可以得到

定理 3.2.1. *设c为常数. 则有*

1. $0 \le Var(X) = EX^2 - (EX)^2$, 因此$Var(X) \le EX^2$.

2. $Var(cX) = c^2 Var(X)$

3. $Var(X)=0$当且仅当$P(X=c)=0$, 其中$c=EX$.

4. 对任何常数c有, $Var(X)\le E(X-c)^2$, 其中等号成立当且仅当$c=EX$.

5. 如果随机变量X和Y相互独立, a,b为常数. 则$Var(aX+bY)=a^2Var(X)+b^2Var(Y)$.

常见分布的方差:

1. 二项分布$X\sim B(n,p)$:

$$VarX=np(1-p)$$

2. Poisson 分布$X\sim P(\lambda)$:

$$VarX=\lambda$$

3. 均匀分布$X\sim U[a,b]$:

$$VarX=\frac{(b-a)^2}{12}$$

4. 指数分布$X\sim Exp(\lambda)$:

$$VarX=1/\lambda^2$$

5. 正态分布$X\sim N(\mu,\sigma^2)$:

$$VarX=\sigma^2$$

由此得到正态分布$N(\mu,\sigma^2)$ 中另一参数σ^2 的解释: 它就是分布的方差, 正态分布完全由其均值μ 和方差σ^2 决定, 故也常称为"均值为μ 方差为σ^2 的正态分布". 方差σ^2 越小, 则X 的取值以更大的概率集中在其均值μ附近.

定义 3.2.2. 我们称

$$X^*=\frac{X-EX}{\sqrt{Var(X)}}$$

为X的标准化随机变量. 易见$EX^*=0, Var(X^*)=1$.

我们引入标准化随机变量是为了消除由于计量单位的不同而给随机变量带来的影响. 例如, 我们考察人的身高, 那么当然可以以米为单位, 得到X_1, 也可以以厘米为单位, 得到X_2. 于是就有得到$X_2=100X_1$. 那么这样一来, X_2与X_1 的分布就有所不同. 这当然是一个不合理的现象. 但是通过标准化, 就可以消除两者之间的差别, 因为我们有$X_2^*=X_1^*$. 对于正态分布, 我们经过标准化$Y=(X-\mu)/\sigma$, 就可以得出均值为0方差为1的正态分布, 即标准正态分布.

§3.2.2 矩

下面我们引入矩的概念，并将之与我们前面所说的期望、方差建立联系.

定义 3.2.3. 设X为随机变量, c 为常数, r为正整数, 则$E[(X-c)^r]$ 称为X 关于c 点的r 阶矩.

比较重要的有两个情况:

1. $c=0$. 这时$\alpha_k=EX^r$ 称为X 的r 阶原点矩.

2. $c=EX$. 这时$\mu_k=E[(X-EX)^r]$ 称为X 的r 阶中心矩.

容易看出, 一阶原点矩就是期望, 二阶中心矩就是X的方差$Var(X)$.

§3.3 协方差和相关系数

现在我们来考虑多维随机向量的数字特征, 以二维的情况为例, 设(X,Y) 为二维随机变量, X, Y本身都是一维随机变量, 那么它们相应的均值方差, 我们都在上两节中讨论过了, 我们更有兴趣的数字特征是反映分量之间关系的那种量, 其中最重要的, 是本节要讨论的协方差和相关系数.

§3.3.1 协方差

定义 3.3.1. 我们称

$$Cov(X,Y)=E(X-EX)(Y-EY)$$

为X与Y的协方差, 其中Cov是英文单词$Covariance$的缩写.

由协方差的定义, 我们立刻可以得到协方差具有如下性质:

1. $Cov(X,Y)=Cov(Y,X)$, $Cov(X,X)=Var(X)$

2. $Cov(X,Y)=EXY-EXEY$, 显然若X、Y 相互独立, 则$Cov(X,Y)=0$

3. $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

4. 对任何实数a_1,a_2,b_1,b_2, 有

$$Cov(a_1X_1+a_2X_2,b_1Y_1+b_2Y_2)=\sum_{i=1}^{2}\sum_{j=1}^{2}a_ib_jCov(X_i,Y_j)$$

§3.3.2 相关系数

定义 3.3.2. 设随机变量X, Y为随机变量, 称

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{VarX} \cdot \sqrt{VarY}},$$

为X与Y的相关系数. 当$\rho_{X,Y} = 0$时, 则称X与Y不相关.

由定义容易看出, 若令$X^* = (X - EX)/\sqrt{VarX}$和$Y^* = (Y - EY)/\sqrt{VarY}$ 分别为X和Y相应的标准化随机变量, 则$\rho_{X,Y} = Cov(X^*, Y^*)$. 因此, 形式上可以把相关系数视为“标准尺度下的协方差”, 从这个角度上说, 相关系数可以更好的反映两个随机变量间的关系, 而不受它们各自所用度量单位的影响.

例 3.3.1. 设$(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$, 则$\rho_{X,Y} = \rho$.

相关系数有如下的性质:

1. 若X和Y相互独立, 则$\rho_{X,Y} = 0$

2. $|\rho_{X,Y}| \le 1$, 等号成立当且仅当X, Y之间存在严格的线性关系, 即

$$\begin{aligned} \rho_{X,Y} &= 1, & \text{则存在 } a > 0, b \in \mathbb{R} \text{ 使得 } X = aY + b \quad & \text{(正相关)} \\ \rho_{X,Y} &= -1, & \text{则存在 } a < 0, b \in \mathbb{R} \text{ 使得 } X = aY + b \quad & \text{(负相关)} \end{aligned}$$

[**注**]: $\rho_{X,Y}$ 也常称作X和Y线性相关系数, 只能刻画X和Y间的线性相依程度, $|\rho_{X,Y}|$ 越接近1, 就表示X, Y 间的线性相关程度越高; $|\rho_{X,Y}| = 0$时, 只是表示X和Y间不存在线性相关, 但可以存在非线性的函数关系.

例 3.3.2. 设$X \sim U(-\frac{1}{2}, \frac{1}{2})$, 而$Y = cosX$, 则

$$Cov(X,Y) = EXY = \int_{-1/2}^{1/2} xcosxdx = 0$$

所以X, Y不相关. 但是X, Y之间存在着非线性的函数关系.

下面我们来讨论不相关与独立性之间的关系.

定理 3.3.1. 对随机变量X, Y, 如果X与Y相互独立, 那么它们一定不相关; 但是如果它们不相关却未必相互独立.

例 3.3.3. 试证明若(X, Y)服从单位圆内的均匀分布, 则X, Y不相关但不独立.

解: 由(X, Y)服从单位圆内的均匀分布, 则(X, Y) 的联合密度函数

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \le 1; \\ 0, & \text{其他}. \end{cases}$$

由此, 可得X和Y的边缘密度函数为

$$f_X(x) = f_Y(x) = \frac{2}{\pi}\sqrt{1 - x^2}, \quad -1 \le x \le 1.$$

因此, $EX = EY = 0$, 又

$$EXY = \int_{-1}^{1} x. \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y. \frac{1}{\pi} dy dx = 0.$$

所以, $Cov(X, Y) = 0$, 从而$\rho_{X,Y} = 0$, 即X和Y不相关. 但由$f(x, y) \neq f_X(x).f_Y(y)$, 知$X$和$Y$显然不独立.

例 3.3.4. 设随机变量X和Y的分布律分别为

$$X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}, \qquad Y \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

并且$P(X \cdot Y = 0) = 1$. 则X与Y不独立, 也不相关.

[注]: 只在正态情形下, 不相关与独立等价. 我们举二维正态的例子来说明, 不妨设$(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$, 则$X$和$Y$独立等价于$\rho = \rho_{X,Y} = 0$, 从而等价于$X$和$Y$不相关.

§3.4 其他一些数字特征与相关函数

- 平均绝对差$E|X - EX|$

表 3.3.1 常见分布表

分布名称	参数	概率密度	期望	方差	特征函数
退化分布	c	$\binom{c}{1}$	c	0	e^{ict}
二点分布	p $(0<p<1)$	$\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$	p	pq	$q+pe^{it}$
二项分布 $B(n,p)$	$n\geq 1$ $0<p<1$	$\binom{n}{k}p^k q^{n-k}$ $k=0,\cdots,n$	np	npq	$(q+pe^{it})^n$
几何分布	p $(0<p<1)$	$q^{k-1}p, k=1,2,\cdots$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1-qe^{it}}$
巴斯卡分布	r,p $r\in\mathbb{N}$ $0<p<1$	$\binom{k-1}{r-1}p^r q^{k-r}$, $k=r,r+1,\cdots$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$(\frac{pe^{it}}{1-qe^{it}})^r$
波松分布$P(\lambda)$	$\lambda(\lambda>0)$	$\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,\cdots$	λ	λ	$e^{\lambda(e^{it}-1)}$
超几何分布	$M,N,n\in\mathbb{N}$	$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$\frac{nM}{N}$	$\frac{nM}{N}\frac{(N-M)}{N}\frac{N-n}{N-1}$	
均匀分布 $U(a,b)$	$a,b(a<b)$	$\frac{1}{b-a}I_{a<x<b}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
正态分布 $N(a,\sigma^2)$	a,σ^2	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$	a	σ^2	$e^{iat-\frac{1}{2}\sigma^2t^2}$
指数分布	$\lambda(\lambda>0)$	$\lambda e^{-\lambda x}I_{x>0}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$(1-\frac{it}{\lambda})^{-1}$
χ^2分布	$n(n\geq 1)$	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$ $x>0$	n	$2n$	$(1-2it)^{-n/2}$

- 矩母函数Ee^{tX}, 其中$t\in\mathbb{R}$.
- 特征函数Ee^{itX}, 其中$t\in\mathbb{R}$, i 为虚数.

定义 3.4.1. 如果离散型随机变量X的分布律为$P(X=a_i)=p_i,\ i\in\mathbb{N}$, 那么

$$Ee^{itX}=\sum_{i=1}^{\infty}e^{ita_i}p_i.$$

如果连续型随机变量X的密度函数为$f(x)$, 那么

$$Ee^{itX}=\int_{-\infty}^{\infty}e^{itx}f(x)dx.$$

§3.5 大数定律和中心极限定理

极限定理是概率论的重要内容,也是数理统计学的基石之一. 中心极限定理,是概率论中讨论随机变量和的分布以正态分布为极限的一组定理,这组定理是数理统计学和误差分析的理论基础, 指出了大量随机变量近似服从正态分布的条件.

§3.5.1 大数定律

定义 3.5.1. 如果对任何$\varepsilon > 0$, 都有

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \varepsilon) = 0,$$

那么我们就称随机变量序列$\{\xi_n,\ n \in \mathbb{N}\}$依概率收敛到随机变量$\xi$, 记为$\xi_n \xrightarrow{p} \xi$.

定理 3.5.1. 设$\{X_n\}$是一列独立同分布*(i.i.d.)*的随机变量序列，具有公共的数学期望μ和方差σ^2. 则

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} \mu,$$

即$\{X_n\}$服从*(*弱*)*大数定律。

[**注**]: 实际上，我们只需要均值存在即有大数定律成立，上述定理中加上了方差存在的条件，只是为了证明的方便。

作为上述定理的一个特例，我们有

例 3.5.1. 如果以ζ_n表示n重*Bernoulli*试验中的成功次数, 则有

$$\frac{\zeta_n}{n} \xrightarrow{p} p.$$

如果用$f_n = \zeta_n/n$表示成功出现的频率,则上例说明$f_n \xrightarrow{p} p$,即频率(依概率)收敛到概率.

为证明定理3.5.1,我们需要如下的Chebyshev 不等式:

引理 3.5.1 (Chebyshev 不等式)**.** 设随机变量X的方差存在，则

$$P(|X - EX| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}, \quad \forall\, \varepsilon > 0.$$

我们可以用Chebyshev不等式来估计X与EX的偏差,但是Chebyshev 不等式作为一个理论工具比作为估计的实际方法要恰当一些, 其重要性在于它的应用普遍性,但是不能希望很普通的命题对一些个别情况给了深刻的结果. 如令X为掷一个均匀的骰子所得到的点数,则$\mu = EX = 7/2$, $\sigma^2 = \mathrm{Var}(X) = 35/12$. X与μ的最大偏差为$2.5 \approx 3\sigma/2$. $|X-\mu|$大于这个偏差的概率为0,然而利用Chebyshev不等式仅仅断定这个概率少于0.47. 这时就需要找更精确的估计.

定理3.5.1的证明. 利用Chebyshev 不等式，并注意到$E\overline{X} = \mu$, $\mathrm{Var}\overline{X} = \sigma^2/n$,我们有,

$$P(|\overline{X} - \mu| \geq \varepsilon) \leq \sigma^2/(n\varepsilon^2) \to 0, \quad n \to \infty, \quad \forall \varepsilon > 0.$$

定理得证.

§3.5.2 中心极限定理

中心极限定理是概率论中讨论随机变量序列的分布收敛于正态分布的一类定理. 它是概率论中最重要的一类定理,有广泛的实际应用背景.在自然界与生产中, 一些随机现象可能会受到许多不确定因素的影响,如果这些彼此之间没有什么依存关系,且谁也没有特别突出的影响,那么,这些影响的累积效应将会使现象近似地服从正态分布.中心极限定理就是从数学上证明了这一现象.

定理 3.5.2. *设$\{X_n\}$为$i.i.d$的随机变量序列，具有公共的数学期望μ 和方差σ^2. 则$X_1 + \cdots + X_n$的标准化形式$\frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - n\mu)$满足中心极限定理. 即对任意$x \in \mathbb{R}$，有*

$$\lim_{n\to\infty} F_n(x) = \Phi(x),$$

其中$F_n(x)$为$\frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - n\mu)$的分布函数, 而$\Phi(x)$为标准正态分布$N(0,1)$的分布函数. 记为

$$\frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - n\mu) \xrightarrow{d} N(0,1).$$

定理3.5.2的令人吃惊之处就是任何独立同分布的随机变量序列,不论它的分布是什么,只要存在有限的方差,那么它们的标准化部分和都渐近于标准正态分布.这也说明了正态分布的普遍性.

由定理3.5.2,我们很容易得到如下推论

定理 3.5.3. *设$X_1, \cdots, X_n$相互独立且具有相同的分布*

$$P(X_1 = 1) = 1 - P(X_1 = 0) = p,\ 0 < p < 1.$$

则有

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

即

$$\lim_{n\to\infty} P\Big(\frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}} \le x\Big) = \Phi(x), \quad \forall\, x \in \mathbb{R}.$$

定理3.5.2称为棣莫弗-拉普拉斯定理,是历史上最早的中心极限定理.因为定理3.5.2中随机变量$X_1, \cdots, X_n$的和$X_1 + \cdots + X_n \sim B(n, p)$, 我们利用正态分布近似地估计二项分布.

设$t_1 < t_2$是两个正整数,则当n相当大时,由定理3.5.2,近似地有

$$P(t_1 \le X_1 + \cdots + X_n \le t_2) \approx \Phi(y_2) - \Phi(y_1),$$

其中

$$y_i = (t_i - np)/\sqrt{np(1-p)},\ i = 1, 2.$$

为提高精度,我们可把y_1, y_2修正为

$$y_1 = (t_1 - 1/2 - np)/\sqrt{np(1-p)}, \quad y_2 = (t_2 + 1/2 - 1/2 - np)/\sqrt{np(1-p)}.$$

例 3.5.2. *设一考生参加100道题的英语标准化考试(每道题均为有四个备选答案的选择题，有且仅有一个答案是正确的)，每道题他都随机地选择一个答案，假设评分标准为:选对得一分，选错或不选不得分。试给出该考生最终得分大于等于25的概率.*

解: 记X_i表示第i题的得分, $i = 1, 2 \cdots, 100$.则$X_1, \cdots, X_n$是一列独立同分布的随机变量具有共同的分布

$$1 - P(X_1 = 0) = P(X_1 = 1) = 0.25.$$

利用中心极限定理,有

$$P(X_1 + \cdots + X_{100} \ge 25) = P\Big(\frac{X_1 + \cdots + X_{100} - 100 * 0.25}{\sqrt{100 * 0.25 * 0.75}} \ge 0\Big) = 1 - \Phi(0) = 1/2.$$

例 3.5.3. 每天有*1000*个旅客需要乘坐火车从芝加哥到洛杉矶, 这两个城市之间有两条竞争的铁路, 它们的火车同时开出同时到达并且具有同样的设备.设这*1000*个人乘坐那一条铁路的火车是相互独立而且又是任意的,于是每列火车的乘客数目可视为概率为$1/2$的1000 重*Bernoulli* 试验中成功的次数.如果一列火车设置$s < n$个座位, 那么一旦有多于s个旅客来乘车就容纳不下了,令这个事件发生的概率为$f(s)$.利用中心极限定理,有

$$f(s) \approx 1 - \Phi\Big(\frac{2s - 1000}{\sqrt{1000}}\Big).$$

要求s使得$f(s) < 0.01$,即在*100*次中有*99*次是有足够的座位的. 查表容易求出$s = 537$.这样,两列火车所有的座位数为*1074*,其中只有*74*个空位,可见由于竞争而带来的损失是很小的.

第四章　数理统计的基本概念及抽样分布

教学目的:

1) 使学生对什么叫数理统计及其发展史有一个初步的了解。

2) 使学生掌握数理统计的若干基本概念，如总体、样本、简单样本、统计模型等。

3) 使学生掌握统计量χ^2、t、F、正态总体样本均值和样本方差的分布及其简单性质.

§4.1　引言

§4.1.1　什么叫数理统计学

本课程的前四章介绍了概率论的基本内容, 为数理统计学建立了重要的数学基础.从本章起, 我们转入本课程的第二部分—数理统计学. 下面我们首先说明什么是数理统计学.

统计学的任务是研究怎样有效地收集、整理和分析带有随机性影响的数据, 从而对所考虑的问题作出一定结论的方法和理论. 它是一门实用性很强的学科, 在人类活动的各个领域有着广泛的应用. 研究统计学方法的理论基础问题的那一部分构成"数理统计学"的内容. 一般地可以认为

数理统计是数学的一个分支, 它是研究如何有效地收集和有效地使用带有随机性影响的数据的一门学科.

下面通过例子对此加以说明.

1. 有效地收集数据

收集数据的方法有: 全面观察(或普查)、抽样调查和安排试验等方式.

例 4.1.1. *人口普查和抽样调查. 我国在2000年进行了第五次人口普查. 如果普查的数据是准确无误的, 无随机性可言, 不需用数理统计方法. 由于人口普查, 调查项目很多, 我国有13亿人口, 普查工作量极大, 而训练有素的工作人员缺乏. 因此虽是全面调查, 但数据并不可靠, 农村超计划生育瞒报、漏报人口的情况时有发生. 针对普查数据不可靠, 国家统计局在人口普查的同时还派出专业人员对全国人口进行抽样调查, 根据抽样调查的结果, 对人口普查的数字进行适当的修正. 抽样调查在普查不可靠时是一种补充办法.*

如何安排抽样调查, 这是有效收集数据的重要问题, 这构成数理统计学的一个重要分支—《抽样调查方法》.

例 4.1.2. *考察某地区10000农户的经济状况. 从中挑选100户做抽样调查. 若该地区分成平原和山区两部分, 平原地区较富, 占该地区农户的70%, 山区的30%农户较穷. 我们的抽样方案规定在抽取的100户中, 从平原地区抽70户, 山区抽30户, 在各自范围内用随机化方法抽取.*

在本例中有效收集数据是通过合理地设计抽样方案来实现的. 在通过试验收集数据的情形如何做到有效收集数据, 请看下例:

例 4.1.3. *某化工产品的得率与温度、压力和原料配方有关. 为提高得率, 通过试验寻找最佳生产条件. 试验因素和水平如下*

因素 \ 样品	*1*	*2*	*3*	*4*
温度	*800*	*1000*	*1200*	*1400*
压力	*10*	*20*	*30*	*40*
配方	*A*	*B*	*C*	*D*

3个因素, 每个因素4个水平共要做 $4^3 = 64$ 次试验. 做这么多试验人力、物力、财力都不可能. 因此, 如何通过尽可能少的试验获得尽可能多的信息? 比如采用正交表安排试验就是一种有效的方法.

如何安排试验方案和分析试验结果, 这构成数理统计的另一分支—《试验的设计和分析》. 在本例中有效收集数据是通过科学安排试验的方法来实现的.

在有效收集数据中一个重要问题是: 数据必须具有随机性.

2. 有效的使用数据

获取数据后, 需要用有效的方法, 去集中和提取数据中的有关信息, 以对所研究的问题作出一定的结论, 在统计上称为"推断".

为了有效的使用数据进行统计推断, 需要对数据建立一个统计模型, 并给定某些准则去评判不同统计推断方法的优劣.

例 4.1.4. *为估计一个物体的重量a,把它在天平上称5次获得数据 $x_1, x_2, \cdots, x_5$, 它们都受到随机性因素的影响(天平的精度反映了影响的大小). 估计a的大小有下列三种不同方法: (1) 用5个数的算术平均值 $\bar{x} = \frac{1}{5}(x_1 + \cdots + x_5)$ 去估计a; (2) 将 $x_1, x_2, \cdots, x_5$ 按大小排列为 $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(5)}$, 取中间一个值 $x_{(3)}$ 去估计a; (3) 用 $W = \frac{1}{2}(x_{(1)} + x_{(5)})$ 去估计a.你可能认为 $\bar{x}$ 优于 $x_{(3)}$, 而 $x_{(3)}$ 优于 W.这是不是对的? 为什么是这样? 在什么条件下才对? 事实上, 对这些问题的研究正是数理统计学的任务.*

要回答这些问题我们需要对数据建立一个统计模型和制定评判不同统计推断方法的准则. 本例中在适当的假定下, 可认为数据服从正态模型.

下面我们举一个例子说明采用合适的统计方法也是有效使用数据的一个重要方面.

例 4.1.5. *某农村有100户农户, 要调查此村农民是否脱贫. 脱贫的标准是每户年均收入超过1万元. 经调查此村90户农户年收入5000元, 10户农户年收入10万元, 问此村农民是否脱贫?*

(1) 用算术平均值计算该村农户年均收入如下:

$$\bar{x} = (90 \times 0.5 + 10 \times 10)/100 = 1.45(\text{万})$$

按此方法得出结论: 该村农民已脱贫. 但90%的农户年均收入只有5000元, 事实上并未脱贫.

(2) 用样本中位数计算该村农户年均收入: 即将100户的年收入记为 x_1, x_2, $\cdots$, x_{100}, 将其按大小排列为 $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(100)}$. 样本中位数定义为排在最中间两户的平均值, 即

$$(x_{(50)} + x_{(51)})/2 = 0.5(\text{万})$$

按此方法得出结论: 该村农民尚未脱贫. 这与实际情况相符.

3. 数理统计方法的归纳性质

数理统计是数学的一个分支, 但是它的推理方法是不一样的. 统计方法的本质是归纳式的, 而数学则是演绎式的. 统计方法的归纳性质, 源于它在作结论时, 是根据所观察到的大量的“个别”情况, “归纳”起来所得. 而不是从一些假设、命题或已知事实出发按一定的逻辑推理得出来的(这后者称为演绎推理) . 举一例子说明: 统计学家通过大量的

观察资料发现, 吸烟与某种呼吸系统的疾病有关. 他得出这一结论的根据是: 从观察到的大量例子, 看到吸烟者中患此种疾病的比例远高于不吸烟者. 他不可能用逻辑推理的方法证明这一点. 试拿统计学与几何学进行比较就可以清楚地看出二者方法的差别所在. 在几何学中要证明"等腰三角形两底角相等", 只需从等腰这个前提出发, 运用几何公理, 一步步地推出这个结论(这一方法属于演绎推理) . 而一个习惯于统计方法的人, 就可能想出这样的方法: 作很多大小形状不一的等腰三角形, 实际测量它的底角查看区别如何, 根据所得数据, 看看可否作出底角相等的结论, 这属于归纳推理的方法.

众所周知, 归纳推理是要冒风险的. 事实上归纳推理的不确定性的出现, 是一种逻辑的必然. 人们不可能做出十分肯定的结论, 因为归纳推理所依据的数据具有随机性. 然而, 不确定性的推理是可行的, 所以推理的不确定性程度是可以计算的. 统计学的作用之一就是提供归纳推理和计算不确定性程度的方法. 不确定性是用概率计算的. 以后会见到我们求参数的**区间估计**, 不但给出**区间估计**的表达式, 而且给出这一估计区间包含未知参数的可靠程度的大小.

§4.1.2 数理统计学的应用

人类在科学研究、生产和管理等各方面的活动, 大都离不开数据资料的收集、整理和分析的工作. 因此统计学的应用领域也及其广泛.

1. 国家行政机关和各种职能机构的工作, 需要经常收集各种有关的数据资料, 以了解情况并做出相应的决策. 这里面的统计工作, 固然有大量的描述性统计的成份, 但统计推断的方法也很有用并且十分必要. 例如在判断某一时期经济运行是否过热, 以便采取宏观调控措施等重大决策时, 对当时经济运行种数据和资料进行定量分析是必不可少的. 这就离不开统计推断方法.

用数理统计方法进行社会调查, 这种工作常属于国家职能部门的工作范围. "抽样调查"是常用的方法. 统计学的方法在决定调查规模和制定有效的抽样方案是很有用, 统计推断方法在对调查得来的资料进行正确分析时也有指导意义. 例如经过精心设计和组织的社会抽样调查, 其效果有时可达到甚至超过全面调查的水平. 在人口学中, 确定一个合适的人口发展动态模型需要掌握大量的观察资料, 而且要使用包括统计方法在内的一些科学方法. 再如, 社会保险基金需要用到精算学, 建立精算模型、对寿命数据的分析都要用到许多统计方法.

2. 在工农业生产中我们常常要利用试验设计和方差分析的方法寻找最佳生产条件. 例如为提高农业中的单位面积产量, 有一些因素对这个指标有影响: 种子的品种、施肥量

和浇水量等; 工业生产中影响某项产品质量指标的因素有原材料产地、配方、温度和压力等因素; 为了找到一组较好的生产条件就要进行试验. 如何科学的安排试验和分析试验结果, 就需要用到统计方法. 试验设计的基本思想和方差分析方法就是R.A. Fisher等在 1923–1926 年期间, 在进行田间试验中发展起来的, 这一方法后来广泛应用于工业生产中.

数理统计方法应用于工业生产的另一个重要方面是产品质量控制、抽样调查和工业产品寿命的可靠性问题. 现代工业生产有批量大和很高可靠度的特点, 需要在连续生产过程中进行工序控制. 成批的产品在交付使用前要进行验收, 这种验收一般不能进行全面检验, 而只能是抽样验收, 需要根据统计学的原理制定合适的抽样方案. 大型设备或复杂产品(如导弹) 包含成千上万个元件. 由于元件的数目很大, 元件的寿命服从一定的概率分布, 整个设备(或产品) 的寿命与其结构和元件的寿命分布有关, 为了估计设备(或产品) 的可靠性, 发展了一系列的统计方法. 统计质量管理就是由上述提到的这些方法构成的.

3. 数理统计方法在经济和金融领域也有广泛的应用, 在经济学中定量分析的趋势比其他社会科学部门更早更深入. 现在有一门叫做"计量经济学"的学科, 其内容主要就是将统计方法(及其他数学方法) 用于分析种种经济问题的数量方面. 例如早在20世纪二、三十年代时间序列的统计分析方法就用于市场预测, 目前在金融等领域也广泛的使用时间序列方法.

4. 统计方法在生物、医学和遗传学中有广泛的应用. 一种药品的疗效如何, 要通过细心安排的试验并使用正确的统计分析方法, 才能比较可靠地做出结论. 分析某种疾病的发生是否与特定因素有关(一个典型的例子是吸烟与患肺癌的关系), 这些问题常常是从观察和分析大量资料的基础上得到启示, 再提高到理论上的研究. 这方面的应用还有流行病数据的统计分析、遗传基因数据的统计分析等.

5. 数理统计方法在气象预报、水文、地震、地质等领域有广泛应用. 在这类领域中, 人们对事物规律性的认识不充分, 使用统计方法有助于获得一些潜在规律性的认识, 用以指导人们的行动.

6. 数理统计方法在科学研究中也具有重要作用. 自然科学研究的根本任务是揭示自然界的规律性, 科学试验是重要手段, 而随机因素对试验结果的影响无所不在. 一个好的统计方法有助于提取观察和实验数据中带根本性的信息, 因而有助于提出较正确的理论或假说. 有了一定的理论和假说后, 统计方法可以指导研究工作者如何进一步安排试验或观察, 以使所得数据更有助于判定定理或假说是否正确. 数理统计学也提供了理

论上有效的方法去估量观察或试验数据与理论的符合程度如何. 一个著名的例子是遗传学中的Mendal定律. 这个根据观察资料提出的定律, 经历了严格的统计检验. 由此可见数理统计方法是科学研究中一个必不可少的手段.

另一方面, 应用上的需要又是统计方法发展的动力. 例如现代统计学的奠基人、英国著名学者R.A. Fisher和K. Pearson在20世纪初期从事统计学的研究, 就是出于生物学、遗传学和农业科学方面的需求.

§4.1.3 统计学发展简史

数理统计学是一门较年轻的学科, 它主要的发展是从20世纪初开始. 大概可分为两个阶段. 前一阶段大致上到第二次世界大战结束时为止. 在这一早期发展阶段中, 起主导作用的是以R.A. Fisher和K. Pearson为首的英国学派, 特别是Fisher, 在本学科的发展中起了独特的作用. 其他一些著名的学者, 如W.S. Gosset (Student)、J. Neyman、E.S. Pearson (K. Pearson的儿子)、A. Wald以及我国的许宝騄教授等都作出了根本性的贡献. 他们的工作奠定了许多统计分支的基础, 提出了一系列具有重要应用价值的统计方法, 和一系列的基本概念和重要理论问题. 有一种意见认为瑞典统计学家H. Cramer在 1946 年发表的著作《Mathematical Methods of Statistics》标志了这门学科达到成熟的地步.

收集和记录种种数据的活动, 在人类历史来源已久. 翻开我国二十四史, 可以看到上面有很多关于钱粮、人口及地震洪水等自然灾害的记录. 在西方国家, Statistics (统计学) 一词源出于State (国家) , 意指国家收集的国情材料. 19世纪中叶以后, 包括政治统计、人口统计、经济统计、犯罪统计、社会统计等多方面内容的"社会统计学"一词在西方开始出现, 与此相应的社会调查也有了较大发展. 人们试图通过社会调查, 搜集、整理、分析数据, 以揭示社会现象和问题, 并提出解决具体问题的方法. 这种情况延续了许多年, 研究方法属于描述统计学的范畴. 这是因为, 没有一定的数学工具特别是概率论的发展, 无法建立现代意义下的数理统计学. 也因为这方面的需求还没达到那么迫切, 足以构成一股强大的推动力. 到十九世纪末和二十世纪初情况才起了较大的变化. 有人认为二十世纪初K. Pearson关于χ^2统计量极限分布的论文可以作为数理统计诞生的一个标志; 也有人认为, 直到1922年Fisher关于统计学的数学基础那篇著名论文的发表, 数理统计才正式诞生.

综上所述, 我们可否可以得到如下粗略的结论: 收集和整理乃至使用观察和试验数据的工作由来已久, 这类活动对于数理统计学的产生, 可算是一个源头. 十九世纪, 特别是十九世纪后半期发展速度加快, 且有了质的变化. 十九世纪未到二十世纪初这一阶段,

出现了一系列的重要工作. 无论如何, 至迟到二十世纪二十年代, 这门科学已稳稳的站住了脚跟. 二十世纪前四十年有了迅速而全面的发展, 到二十世纪四十年代时, 已形成为一个成熟的数学分支.

从战后到现在可以说是第二阶段. 在这个时期中, 许多战前开始形成的数理统计分支, 在战后得到纵深的发展, 理论上的深度也比以前大大加强了. 同时还出现了带根本性的发展, 如Wald的统计判决理论和**Bayes学派**的兴起. 在数理统计的应用方面, 也给人印象深刻. 这不仅是战后工农业生产和科学技术迅速发展所提出的要求, 也是由于电子计算机这一有力工具的出现和飞速发展推动了数理统计学的进步. 战前由于计算工具跟不上, 许多需要大量计算的统计方法很难得以使用. 战后有了高速计算机便变得很容易, 这就大大推广了统计方法的应用. 目前, 统计方法仍在蓬勃发展中. 在一些统计学发达的国家中, 特别在美国, 这方面的人才数以十万计, 并在大多数大学中建立了统计系. 近三十年来数理统计学在我国的发展也是令人瞩目的.

§4.2 数理统计的若干基本概念

§4.2.1 总体和样本

通过下面的例子说明总体、个体和样本的概念.

例 4.2.1. *假定一批产品有10000件, 其中有正品也有废品, 为估计废品率, 我们往往从中抽取一部分, 如100件进行检查. 此时这批10000件产品称为***总体***, 其中的每件产品称为***个体***, 而从中抽取的100件产品称为***样本***. 样本中个体的数目称为***样本的大小***, 也称为***样本容量***. 而抽取样本的行为称为***抽样***.*

从本例我们可对总体和样本作如下直观的定义:

总体是与我们所研究的问题有关的所有个体组成, 而**样本**是总体中抽取的一部分个体.

若总体中个体的数目为有限个, 则称为**有限总体**, 否则称为**无限总体**.

在统计研究中, 人们所关心的不是总体内个体的本身, 而是关心个体上的一项(或几项) 数量指标, 如日光灯的寿命, 零件的尺寸. 在例4.2.1中若产品为正品用0表示, 若产品为废品用1表示, 我们关心的个体取值是0还是1. 因此我又可获得总体的如下定义:

总体可以看成有所有个体上的某种**数量指标**构成的集合, 因此它是**数**的集合.

由于每个个体的出现是随机的, 所以相应的个体上的数量指标的出现也带有随机性. 从而可以把此种数量指标看成随机变量, 随机变量的分布就是该数量指标在总体中的分布. 以例4.2.1来说明, 假定10000只产品中废品数为100件, 其余的为正品, 废品率为0.01. 我们定义随机变量X如下:

$$X = \begin{cases} 1 & \text{废品} \\ 0 & \text{正品}, \end{cases}$$

其概率分布为0–1分布,且有$P(X = 1) = 0.01$. 因此,特定个体上的数量指标是随机变量X的观察值. 这样一来, 总体可以用一个随机变量X 及其分布来描述, 获得如下定义:

定义 4.2.1. *一个统计问题所研究的对象的全体称为总体.在数理统计学中总体可以用一个随机变量及其概率分布来描述.*

由于总体的特征由其分布来刻画, 因此统计学上常把总体和总体分布视为同义语. 由于这个缘故, 常用随机变量的符号或分布的符号来表示总体. 比如研究某批日光灯寿命时, 人们关心的数量指标是寿命X, 那么此总体就可以用随机变量X来表示, 或用其分布函数F 来表示. 若F有密度, 记为f, 则此总体也可用密度函数f来表示. 有时也根据总体分布的类型来称呼总体的名称, 如正态总体、二项分布总体、0–1分布总体. 若总体分布函数记为F, 当有一个从该总体中抽取的相互独立同分布(i.i.d.)的大小为n的样本$X_1, \cdots, X_n$,则常记为

$$X_1, \cdots, X_n \text{ i.i.d. } \sim F \tag{4.2.1}$$

若F有密度f,可记为

$$X_1, \cdots, X_n \text{ i.i.d. } \sim f \tag{4.2.2}$$

若所考虑的总体用随机变量X表示其分布函数为F,则样本$X_1, \cdots, X_n$ 可视为随机变量X的观察值, 亦可记为

$$X_1, \cdots, X_n \text{ i.i.d. } \sim X \tag{4.2.3}$$

(4.2.1)、(4.2.1)和(4.2.3)表示相同的意思.

当个体上的数量指标不止一项时, 我们用随机向量来表示总体. 例如研究某地区小学生的发育状况时, 人们关心的是其身高X和体重Y这两个数量指标, 此时总体就可以用二维随机向量(X, Y)或其联合分布$F(x, y)$表示.

§4.2.2 样本的两重性和简单随机样本

1、样本的两重性

当我们从总体中作具体抽样时, 每次抽样的结果都是些具体的数, 如例5.2.3的打靶问题中, 3维样本$X = (X_1, X_2, X_3)$,其中$0 \leq X_i \leq 10$为整数, $i = 1, 2, 3$,它是数字向量. 但若是在相同条件下, 再打三发, 由于种种不可控制的随机因素的影响, 中靶的环数不可能和上一次完全一样, 具有随机性. 如果无穷次打下去, 每次打三发, 出现的结果可视为随机向量(X_1, X_2, X_3)的观察值.

样本的两重性是说, 样本既可看成具体的数, 又可以看成随机变量(或随机向量). 在完成抽样后，它是具体的数；在实施抽样前，它被看成随机变量. 因为在实施具体抽样之前无法预料抽样的结果, 只能预料它可能取值的范围,故可把它看成一个随机变量，因此才有概率分布可言。为区别起见, 今后用大写的英文字母表示随机变量或随机向量, 用小写字母表示具体的观察值.

对理论工作者, 更重视样本是随机变量这一点, 而对应用工作者虽则将样本看成具体的数字, 但仍不可忽视样本是随机变量(或随机向量) 这一背景. 否则, 样本就是一堆杂乱无章毫无规律可言的数字, 无法进行任何统计处理. 样本既然是随机变量(或随机向量), 就有分布而言, 这样才存在统计推断问题.

2、简单随机样本

抽样是指从总体中按一定方式抽取样本的行为. 抽样的目的是通过取得的样本对总体分布中的某些未知因素做出推断, 为了使抽取的样本能很好的反映总体的信息, 必须考虑抽样方法. 最常用的一种抽样方法叫作"简单随机抽样", 它要求满足下列两条:

(1) 代表性. 总体中的每一个体都有同等机会被抽入样本, 这意味着样本中每个个体与所考察的总体具有相同分布. 因此, 任一样本中的个体都具有代表性.

(2) 独立性. 样本中每一个体取什么值并不影响其它个体取什么值. 这意味着, 样本中各个体$X_1, X_2, \cdots, X_n$ 是相互独立的随机变量.

由简单随机抽样获得的样本$(X_1, \cdots, X_n)$称为简单随机样本. 用数学语言将这一定义叙述如下:

定义 4.2.2. *设有一总体F, $X_1, \cdots, X_n$为从F中抽取的容量为n的样本, 若*

(i) $X_1, \cdots, X_n$*相互独立,*

(ii) $X_1, \cdots, X_n$相同分布, 即同有分布F,
则称$(X_1, \cdots, X_n)$为简单随机样本, 有时简称简单样本或随机样本.

设总体为F, $(X_1, \cdots, X_n)$为从此总体中抽取的简单样本, 则$X_1, \cdots, X_n$的联合分布为:

$$F(x_1) \cdot F(x_2) \cdot \cdots F(x_n) = \prod_{i=1}^{n} F(x_i)$$

若F有密度f,则其联合密度为

$$f(x_1) \cdot f(x_2) \cdot \cdots f(x_n) = \prod_{i=1}^{n} f(x_i)$$

若样本是多维的, 例如从一大群人中抽取n个人, 每个测出其身高和体重. 用随机向量(X, Y) 或其分布$F(x, y)$记总体, $(X_1, Y_1), \cdots, (X_n, Y_n)$就是从这一总体中抽取的一组样本空间, 其联合分布为

$$F(x_1, y_1) \cdot F(x_2, y_2) \cdot \cdots F(x_n, y_n) = \prod_{i=1}^{n} F(x_i, y_i)$$

若$F(x, y)$有密度$f(x, y)$,则其联合密度为

$$f(x_1, y_1) \cdot f(x_2, y_2) \cdot \cdots f(x_n, y_n) = \prod_{i=1}^{n} f(x_i, y_i)$$

显然，有放回抽样获得的样本是简单样本. 当总体中个体数较大或所抽样本在总体中所占比例较小时, 无放回抽样获得的样本可以近似认为是简单样本.

§4.2.3 统计模型

所谓一个问题的**统计模型**, 就是指研究该问题时所抽样本的**样本分布**, 也常称为**概率模型**或**数学模型**.

由于模型只取决于样本的分布, 故常把分布的名称作为模型的名称.如下列例4.2.2中样本分布为正态，可称其为正态模型。因此把模型和样本紧密联系起来是必要的. 统计分析的依据是样本, 从统计上说, 只有规定了样本的分布, 问题才算真正明确了.

下例告诉我们是怎样由一个具体问题建立统计模型的.

例 4.2.2. *为估计一物件的重量a, 用一架天平将它重复称n次, 结果记为$X_1, \cdots, X_n$, 求样本$X_1, \cdots, X_n$的联合分布.*

解: 要定出$X_1, \cdots, X_n$的分布, 就没有前面例子那种简单的算法, 需作一些假定: (1) 假定各次称重是独立进行的, 即某次称重结果不受其它次称重结果的影响. 这样$X_1, \cdots, X_n$就可以认为是相互独立的随机变量. (2) 假定各次称重是在"相同条件"下进行的, 可理解为每次用同一天平, 每次称重由同一人操作, 且周围环境(如温度、湿度等) 都相同. 在这个假定下, 可认为$X_1, \cdots, X_n$是同分布的. 在上述两个假定下, $X_1, \cdots, X_n$是n个独立同分布的随机变量, 即为**简单随机样本**.

为确定$X_1, \cdots, X_n$的联合分布, 在以上假定之下求出X_1的分布即可. 在此考虑称重误差的特性: 这种误差一般由大量的、彼此独立起作用的随机误差迭加而成, 而每一个起的作用都很小. 由概率论中的中心极限定理可知这种误差近似服从正态分布. 再假定天平没有系统误差, 则可进一步假定此误差为均值为0的正态分布. 可以把X_1 (它可视为物重a加上称量误差之和)的概率分布为$N(a, \sigma^2)$.因此简单随机样本$X_1, \cdots, X_n$的联合分布为

$$f(x_1, \cdots, x_n) = (\sqrt{2\pi}\sigma)^{-n} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\} \tag{4.2.4}$$

本例中求样本分布,引入两种假定: (i)导出样本$X_1, \cdots, X_n$ i.i.d.的假定, (ii)正态假定, 这一点依据问题的性质、概率论的极限理论和以往经验.

在有了研究统计模型后, 很多性质不一样的问题, 可以归入到同一模型下. 例如涉及到测量误差的问题, 只要例4.2.2中叙述的假定误差服从正态分布的理由成立, 则都可以用正态模型(4.2.4). 只要把这个模型中的统计问题研究清楚了, 就可以解决许多不同专业部门中的这样一类问题.

另一方面，同一模型下可以提出很多不同的统计问题. 如例4.2.2的$N(a, \sigma^2)$模型中, 有了样本$X_1, \cdots, X_n$, 并规定分布(4.2.4)后就有了一个统计模型. 在这个模型下可提出一些统计问题, 如在例4.2.2中, 我们的问题是估计物重a. 为了考察天平的精度我们可以提出估计σ^2的问题, 当然我们还可以对a和σ^2提出**假设检验**和**区间估计**问题等等.

§4.2.4 统计推断

从总体中抽取一定大小的样本去推断总体的概率分布的方法称为**统计推断**.

数理统计是着手于样本, 着眼于总体, 其任务是用样本去推断总体. 当样本分布完全已知时是不存在任何统计推断问题.

当样本的分布形式已知, 但含有未知参数时, 有关其参数的推断, 称为**参数统计推断**.

在另一些问题中, 情况就要复杂一些. 这类问题中样本分布的形式完全未知, 有关其分布的统计推断问题称为**非参数统计推断**问题.

参数统计推断有种种不同的形式: 主要有**参数估计**和**假设检验**问题. 如例4.2.2中样本分布(亦即总体分布) $N(a,\sigma^2)$ 中, 当a和σ^2未知时, 从总体中抽取大小为n 的样本X_1, $\cdots$, X_n, 对a和σ^2的取值作出估计,或对断言"$a\le 1$"作出接受或拒绝这一假设的结论.

非参数问题中,统计推断的主要任务是通过样本对总体分布的形式作出推断.

由于样本的随机性, 统计推断的结论不可能100%的正确, 但我们可以给出衡量推断正确程度的指标. 如在例4.2.2中, 若用$\bar X=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$估计$a$,可以算出$\bar X$与$a$ 的偏差大于c的概率, 即$P(|\bar X-a|>c)$, 作为用$\bar X$推断a的正确性的合理指标.

统计推断包括下列三方面内容: (1) 提出种种的统计推断的方法. (2) 计算有关统计推断方法性能的数量指标, 如前述例子中用$\bar X$ 估计$N(a,\sigma^2)$中的a,用$P(|\bar X-a|>c)$表示推断性能的数量指标. (3) 在一定的条件和优良性准则下寻找最优的统计推断方法, 或证明某种统计推断方法是最优的.

§4.3 统计量

§4.3.1 统计量的定义

数理统计的任务是通过样本去推断总体. 而样本自身是一些杂乱无章的数字, 要对这些数字进行加工整理, 计算出一些有用的量, 这就如同为了织布, 我们首先要把棉花加工纺成纱, 然后利用纱去织布. 可以这样理解: 这种由样本算出来的量, 把样本中与所要解决的问题有关的信息集中起来了. 我们把这种量称为统计量, 其定义如下:

定义 4.3.1. *由样本算出的量是***统计量***, 或曰,***统计量***是样本的函数.*

对这一定义我们作如下几点说明:

(1) 统计量只与样本有关, 不能与未知参数有关. 例如$X\sim N(a,\sigma^2)$, $X_1,\cdots,X_n$是从总体X中抽取的i.i.d.样本, 则$\sum\limits_{i=1}^{n}X_i$和$\sum\limits_{i=1}^{n}X_i^2$都是统计量, 当$a$和$\sigma^2$皆为未知参数时, $\sum\limits_{i=1}^{n}(X_i-a)$和$\sum\limits_{i=1}^{n}X_i^2/\sigma^2$都不是统计量.

(2) 由于样本具有两重性, 即样本既可以看成具体的数, 又可以看成随机变量; 统计量是样本的函数, 因此统计量也具有两重性. 正因为统计量可视为随机变量(或随机向量), 因此才有概率分布可言, 这是我们利用统计量进行统计推断的依据.

(3) 在什么问题中选用什么统计量, 要看问题的性质. 一般说来, 所提出的统计量应是最好的集中了样本中与所讨论问题有关的信息, 这不是容易做到的.

§4.3.2 若干常用的统计量

1. 样本均值: 设$X_1, \cdots, X_n$是从某总体X中抽取的样本,则称

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

为**样本均值**. 它分别反映了总体均值的信息.

2. 样本方差: 设$X_1, \cdots, X_n$是从某总体X中抽取的样本,则称

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

为**样本方差**,它分别反映总体方差的信息.而S称为样本标准差,它反映了总体标准差的信息.

3. 样本矩: 设$X_1, \cdots, X_n$为从总体F中抽取的样本, 则称

$$a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k = 1, 2, \cdots$$

为**样本k阶原点矩**, 特别$k=1$时, $a_1 = \bar{X}$即**样本均值**. 称

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k, \quad k = 2, 3, \cdots$$

为**样本k阶中心矩**.

4. 次序统计量及其有关统计量: 设$X_1, \cdots, X_n$为从总体F中抽取的样本, 把其按大小排列为$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$,则称$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$为**次序统计量**, $(X_{(1)}, \cdots, X_{(n)})$的任一部分也称为**次序统计量**.

利用次序统计量可以定义下列统计量:

(1) **样本中位数**:

$$m_{\frac{1}{2}} = \begin{cases} X_{(\frac{n+1}{2})} & \text{当n为奇数} \\ \frac{1}{2}[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}] & \text{当n为偶数} \end{cases} \tag{4.3.1}$$

样本中位数反映总体中位数的信息. 当总体分布关于某点对称时, 对称中心既是总体中位数又是总体均值, 故此时$m_{1/2}$也反映总体均值的信息.

(2) **极值**: $X_{(1)}$和$X_{(n)}$称为样本的极小值和极大值. 极值统计量在关于灾害问题和材料试验的统计分析中是常用的统计量.

§4.4 三大分布—χ^2, t, F分布及正态总体样本均值和样本方差的分布

能求出抽样分布的确切而且具有简单表达式的情形并不多,一般都较难. 所幸的是,在总体分布为正态情形, 许多重要统计量的抽样分布可以求得,这些多与下面讨论的三种分布有密切关系. 这三个分布在后面几章中有重要应用.

§4.4.1 χ^2分布

定义 4.4.1. 设$X_1, X_2, \cdots, X_n$ $i.i.d. \sim N(0,1)$,令$X = \sum_{i=1}^{n} X_i^2$,则称X是自由度为n的χ^2变量, 其分布称为自由度为n的χ^2**分布**, 记为$X \sim \chi_n^2$.

设随机变量X是自由度为n的χ^2随机变量, 则其概率密度函数为

$$g_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0. \end{cases} \tag{4.4.1}$$

注 4.4.1. 若记$\Gamma(\alpha, \lambda)$表示形状参数为α、刻度参数为λ的*Gamma*分布, 其密度函数如下

$$p(x;\alpha,\lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

则自由度为n的χ^2**分布**与***Gamma***分布的关系为: $X = \sum_{i=1}^{n} X_i^2 \sim \Gamma(n/2, 1/2)$. 我们也可以利用这一关系给出$\chi^2$分布的定义: "若随机变量$X$的概率密度函数为$\Gamma(n/2, 1/2)$, 则称$X$为服从自由度为$n$的$\chi^2$分布".

χ_n^2的密度函数$g_n(x)$形状如图4.4.1 .

χ_n^2密度函数的支撑集(即使密度函数为正的自变量的集合)为$(0, +\infty)$, 由图4.4.1可见当自由度n越大, χ_n^2 的密度曲线越趋于对称, n越小, 曲线越不对称. 当$n = 1, 2$时曲线是单调下降趋于0. 当$n \ge 3$时曲线有单峰, 从0开始先单调上升, 在一定位置达到峰值, 然后单下降趋向于0.

若$X \sim \chi_n^2$,记$P(X > c) = \alpha$, 则$c = \chi_n^2(\alpha)$称为χ_n^2分布的上侧α分位数,如图4.4.2所示. 当α和n给定时可查表求出$\chi_n^2(\alpha)$之值, 如$\chi_{10}^2(0.01) = 23.209$, $\chi_5^2(0.05) = 12.592$等.

χ^2变量具有下列性质:

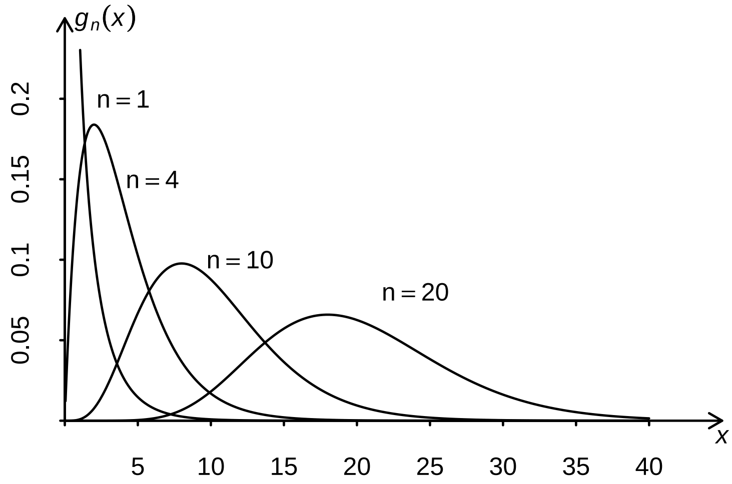


图 4.4.1 χ^2_n的密度函数$g_n(x)$形状图

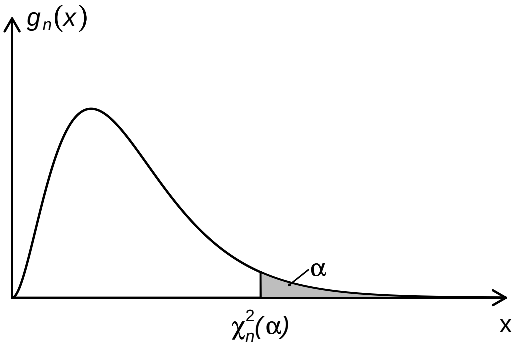


图 4.4.2 χ^2_n的上侧α分位数

(1) 设随机变量$X \sim \chi^2_n$则有$E(X) = n,\ Var(X) = 2n.$

(2) 设$Z_1 \sim \chi^2_{n_1},\ Z_2 \sim \chi^2_{n_2}$,且$Z_1$和$Z_2$独立, 则$Z_1 + Z_2 \sim \chi^2_{n_1+n_2}$.

我们从X^2分布的定义出发给出一个简单证明: 由定义$Z_1 = X_1^2 + \cdots + X_{n_1}^2$, 此处

$$X_1, X_2, \cdots, X_{n_1} \text{ i.i.d.} \sim N(0,1),$$

同理$Z_2 = X_{n_1+1}^2 + \cdots + X_{n_1+n_2}^2$, 此处

$$X_{n_1+1}, X_{n_1+2}, \cdots, X_{n_1+n_2} \text{ i.i.d.} \sim N(0,1),$$

再由Z_1和Z_2的独立性可知

$$X_1, X_2, \cdots, X_{n_1}, X_{n_1+1}, \cdots, X_{n_1+n_2} \text{ i.i.d.} \sim N(0,1).$$

因此

$$Z_1+Z_2=X_1^2+\cdots+X_{n_1}^2+X_{n_1+1}^2+\cdots+X_{n_1+n_2}^2.$$

按定义即有$Z_1+Z_2\sim\chi^2_{n_1+n_2}$.

§4.4.2　t分布

定义 4.4.2. 设随机变量$X\sim N(0,1), Y\sim\chi_n^2$, 且$X$和$Y$独立, 则称

$$T=\frac{X}{\sqrt{Y/n}}$$

为自由度为n的t变量, 其分布称为有自由度n的t**分布**,记为$T\sim t_n$.

设随机变量$T\sim t_n$,则其密度函数为

$$t_n(x)=\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\quad -\infty<x<\infty \tag{4.4.2}$$

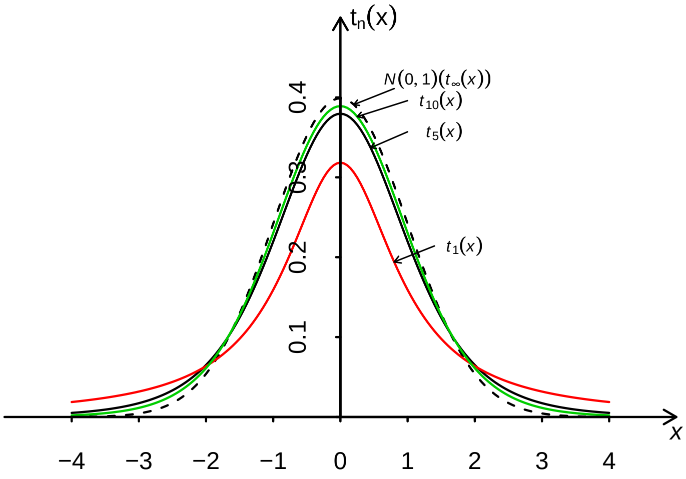


图 4.4.3　t_n的密度函数$t_n(x)$形状图

t_n的密度函数与标准正态分布$N(0,1)$密度很相似, 它们都是关于原点对称, 单峰偶函数, 在$x=0$处达到极大. 但t_n的峰值低于$N(0,1)$的峰值, t_n的密度函数尾部都要比$N(0,1)$的两侧尾部粗一些. 如图4.4.3所示. 容易证明: $\lim\limits_{n\to\infty}t_n(x)=\varphi(x)$,此处$\varphi(x)$是$N(0,1)$变量的密度函数.

若$T\sim t_n$,记$P(|T|>c)=\alpha$,则$c=t_n(\alpha/2)$为自由度为n的t分布的双侧α分位数(如图4.4.4所示). 当给定α时, $t_n(\alpha)$, $t_n(\alpha/2)$等可通过查表求出. 例如$t_{12}(0.05)=1.782$, $t_9(0.025)=2.262$等.

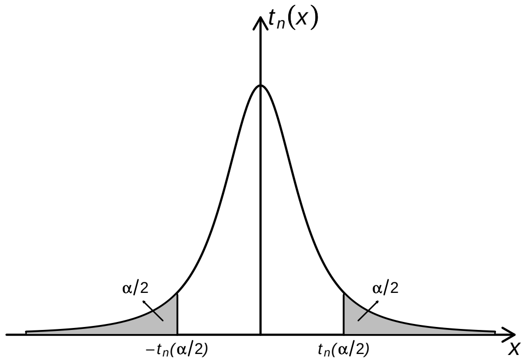


图 4.4.4 t_n的双侧α分位数

t分布是英国统计学家W.S. Gosset 在1908年以笔名Student 发表的论文中提出的, 故后人称为"学生氏(Student)分布" 或"t**分布**".

t变量具有下列的性质:

(1)若随机变量$T \sim t_n$,则当$n \geq 2$时, $E(T) = 0$. 当$n \geq 3$时, $Var(T) = \frac{n}{n-2}$.

(2)当$n \to \infty$时, t变量的极限分布为$N(0, 1)$.

§4.4.3 F分布

定义 4.4.3. 设随机变量$X \sim \chi^2_m$, $Y \sim \chi^2_n$,且X和Y独立,则称

$$F = \frac{X/m}{Y/n}$$

为自由度分别是m和n的F变量, 其分布称为自由度分别是m和n的F**分布**, 记为$F \sim F_{m,n}$.

若随机变量$Z \sim F_{m,n}$, 则其密度函数为

$$f_{m,n}(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1} (n + mx)^{-\frac{m+n}{2}}, & x > 0, \\ 0, & \text{其它}. \end{cases} \tag{4.4.3}$$

自由度为m, n的F分布的密度函数如图4.4.5 . 注意F分布的自由度m和n是有顺序的, 当$m \neq n$时若将自由度m和n的顺序颠倒一下, 得到的是两个不同的F分布. 图4.4.5中给出了几个不同自由度的密度函数的曲线. 由图4.4.5 可见对给定$m = 10$, n取不同值时$f_{m,n}(x)$的形状, 我们看到曲线是偏态的, n越小偏态越严重.

若$F \sim F_{m,n}$,记$P(F > c) = \alpha$, 则$c = F_{m,n}(\alpha)$称为F分布的上侧α分位数(见图4.4.6). 当m, n和α给定时, 可以通过查表求出$F_{m,n}(\alpha)$之值, 例如$F_{4,10}(0.05) = 3.48$, $F_{10,15}(0.01) = 3.80$等. 这在区间估计和假设检验问题中常常用到.

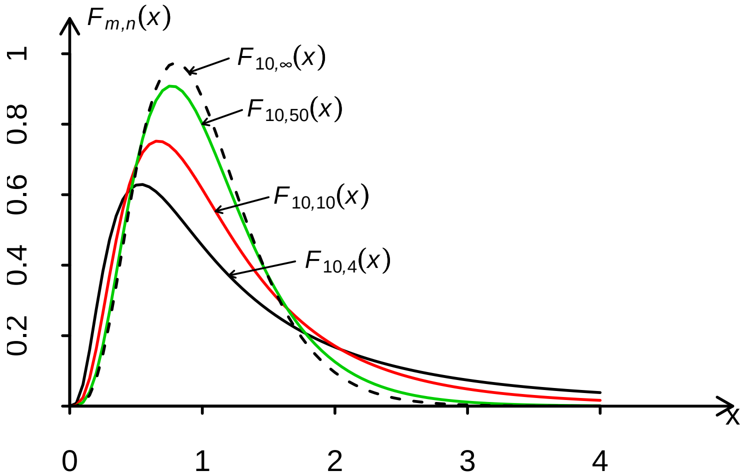


图 4.4.5 $F_{m,n}$的密度函数$f_{m,n}(x)$形状图

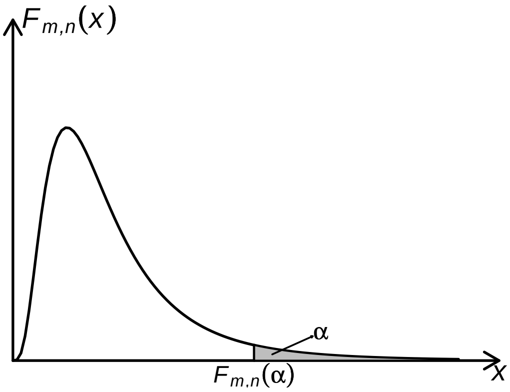


图 4.4.6 $F_{m,n}$的上侧α分位数

F变量具有下列的性质:

(1) 若$Z \sim F_{m,n}$,则$1/Z \sim F_{n,m}$.

(2) 若$T \sim t_n$,则$T^2 \sim F_{1,n}$

(3) $F_{m,n}(1-\alpha) = 1/F_{n,m}(\alpha)$

以上性质中(1)和(2)是显然的, (3)的证明不难,留给读者作为练习. 尤其性质(3)在求区间估计和假设检验问题时会常常用到. 因为当α为较小的数,如$\alpha = 0.05$或$\alpha = 0.01$, m, n给定时,从已有的F分布表上查不到$F_{m,n}(1-0.05)$和$F_{m,n}(1-0.01)$之值, 但它们的值可利用性质(3)求得, 因为$F_{n,m}(0.05)$和$F_{n,m}(0.01)$ 是可以通过查F分布表求得的.

§4.4.4 正态总体样本均值和样本方差的分布

为方便讨论正态总体样本均值和样本方差的分布, 我们先给出正态随机变量的线性函数的分布.

1. 正态变量线性函数的分布

设随机变量$X_1, \cdots, X_n$ $i.i.d. \sim N(a, \sigma^2)$, $c_1, c_2, \cdots, c_n$ 为常数, 则有

$$T = \sum_{k=1}^{n} c_k X_k \ \sim \ N\left(a \sum_{k=1}^{n} c_k, \sigma^2 \sum_{k=1}^{n} c_k^2\right)$$

特别,当$c_1 = \cdots = c_n = 1/n$,即$T = \frac{1}{n} \sum\limits_{i=1}^{n} X_i = \bar{X}$时,有

$$\bar{X} \sim N(a, \sigma^2/n).$$

2. 正态变量样本均值和样本方差的分布

下述定理给出了正态变量样本均值和样本方差的分布和它们的独立性.

定理 4.4.1. *设$X_1, X_2, \cdots, X_n$ $i.i.d. \sim N(a, \sigma^2), \bar{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$和$S^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \bar{X})^2$分别为样本均值和样本方差,则有*

(1) $\bar{X} \sim N(a, \frac{1}{n}\sigma^2)$;

(2) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$;

(3) $\bar{X}$*和*S^2*独立.*

定理的证明超出我们的要求, 只要求记住这一结论.

§4.4.5 几个重要推论

下面几个推论在正态总体区间估计和假设检验问题中有着重要应用.

推论 4.4.1. *设$X_1, X_2, \cdots, X_n$相互独立相同分布(i.i.d.) $\sim N(a, \sigma^2)$, 则*

$$T = \frac{\sqrt{n}(\bar{X} - a)}{S} \sim t_{n-1}.$$

证: 由注5.4.3可知$\bar{X} \sim N(a, \sigma^2/n)$, 将其标准化得$\sqrt{n}(\bar{X}-a)/\sigma \sim N(0,1)$. 又$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, 即$S^2/\sigma^2 \sim \chi^2_{n-1}/(n-1)$, 且$\bar{X}$和$S^2$独立, 按定义有

$$T = \frac{\sqrt{n}(\bar{X}-a)/\sigma}{\sqrt{S^2/\sigma^2}} = \frac{\sqrt{n}(\bar{X}-a)}{S} \sim t_{n-1}.$$

推论 4.4.2. *设$X_1, X_2, \cdots, X_m$ $i.i.d. \sim N(a_1, \sigma_1^2)$, $Y_1, Y_2, \cdots, Y_n$ $i.i.d. \sim N(a_2, \sigma_2^2)$,且假定$\sigma_1^2 = \sigma_2^2 = \sigma^2$, 样本$X_1, X_2, \cdots, X_m$ 与$Y_1, Y_2, \cdots, Y_n$独立, 则*

$$T = \frac{(\bar{X}-\bar{Y})-(a_1-a_2)}{S_w} \cdot \sqrt{\frac{mn}{n+m}} \sim t_{n+m-2},$$

此处$(n+m-2)S_w^2 = (m-1)S_1^2 + (n-1)S_2^2$, 其中

$$S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i-\bar{X})^2, \qquad S_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j-\bar{Y})^2.$$

证: 由注5.4.3可知$\bar{X} \sim N(a, \sigma^2/m), \bar{Y} \sim N(a_2, \sigma^2/n)$, 故有$\bar{X}-\bar{Y} \sim N\big(a_1-a_2, (\frac{1}{m}+\frac{1}{n})\sigma^2\big) = N\big(a_1-a_2, \frac{n+m}{mn}\sigma^2\big)$. 将其标准化得

$$\frac{\bar{X}-\bar{Y}-(a_1-a_2)}{\sigma}\sqrt{\frac{mn}{m+n}} \sim N(0,1). \tag{4.4.4}$$

又$(m-1)S_1^2/\sigma^2 \sim \chi^2_{m-1}$, $(n-1)S_2^2/\sigma^2 \sim \chi^2_{n-1}$, 再利用$\chi^2$分布的性质可知

$$\frac{(m-1)S_1^2+(n-1)S_2^2}{\sigma^2} \sim \chi^2_{n+m-2}. \tag{4.4.5}$$

再由(4.4.4)和(4.4.5)中$(\bar{X}, \bar{Y})$ 与(S_1^2, S_2^2)相互独立, 由定义可知

$$\begin{aligned} T &= \frac{(\bar{X}-\bar{Y})-(a_1-a_2)}{\sigma}\sqrt{\frac{mn}{n+m}}\Bigg/\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{\sigma^2(n+m-2)}} \\ &= \frac{(\bar{X}-\bar{Y})-(a_1-a_2)}{S_w}\sqrt{\frac{nm}{n+m}} \ \sim \ t_{n+m-2}. \end{aligned}$$

推论 4.4.3. *设$X_1, X_2, \cdots, X_m$ $i.i.d. \sim N(a_1, \sigma_1^2)$, $Y_1, Y_2, \cdots, Y_n$ $i.i.d. \sim N(a_2, \sigma_2^2)$, 且合样本$X_1, X_2, \cdots, X_m$和$Y_1, Y_2, \cdots, Y_n$相互独立, 则*

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{m-1,n-1},$$

此处S_1^2和S_2^2定义如推论4.4.2所述.

证: 由注5.4.3可知$(m-1)S_X^2/\sigma_1^2 \sim \chi^2_{m-1}$, $(n-1)S_Y^2/\sigma_2^2 \sim \chi^2_{n-1}$, 且二者独立, 由$F$分布的定义可知

$$F = \frac{\frac{(m-1)S_X^2}{\sigma_1^2}/(m-1)}{\frac{(n-1)S_Y^2}{\sigma_2^2}/(n-1)} = \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{m-1,n-1}.$$

证毕.

第五章　参数估计

教学目的:

1) 让学生理解矩估计和极大似然估计方法.

2) 理解置信区间定义.

3) 掌握常见的总体分布下参数的点估计和置信区间的计算.

设有一个总体, 以$f(x;\theta_1,\cdots,\theta_k)$记其概率密度函数(若总体分布是连续性的), 或其概率函数(若总体分布为离散型的). 为叙述方便我们统一称$f(x;\theta_1,\cdots,\theta_k)$为总体的概率函数. 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者参数的某些函数. 一般假定总体分布形式已知，未知的仅仅是一个或几个参数. 利用从总体$f(x;\theta_1,\cdots,\theta_k)$中抽取的一组样本$X_1,\cdots,X_n$去对参数$\theta_1,\cdots,\theta_k$的未知值作出估计或估计它们的某个已知函数$g(\theta_1,\cdots,\theta_k)$.

§5.1　点估计

设总体X的分布函数形式已知, 但它的一个或多个参数为未知, 例如参数θ未知, 根据样本$X_1,\cdots,X_n$来估计参数θ, 就是要构造适当的统计量$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$. 当有了样本$X_1,\cdots,X_n$的值后,就代入$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$中算出一个值,用来作为$\theta$的估计值. 为这样特定目的而构造的统计量$\hat{\theta}$叫做$\theta$的**估计量**. 由于参数$\theta$是数轴上的一个点, 用$\hat{\theta}$估计$\theta$,等于用一个点去估计另一个点, 所以这样的估计叫做**点估计**.

求点估计的方法有多种, 下面介绍两种点估计方法:

§5.1.1　矩估计方法

矩方法追溯到19世纪的Karl Pearson. 矩方法是基于一种简单的“替换”思想建立起来的一种估计方法. 其基本思想是用样本矩估计总体矩. 由大数律，如果未知参数和总体的某个(些)矩有关系，我们很自然的来构造未知参数的估计。

回忆一下以前关于矩的记法：

$$\text{样本}k\text{阶矩:}\quad a_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \qquad m_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$$

$$\text{总体}k\text{阶矩}: \quad \alpha_k = EX^k \qquad \mu_k = E(X-EX)^2$$

因此在k阶矩存在的情况下，根据大数律有

$$a_k \xrightarrow{p} \alpha_k, \quad m_k \xrightarrow{p} \mu_k$$

从而我们可以使用a_k, m_k分别估计α_k, μ_k。介绍如下: 假设总体X包含k个未知参数$\theta_1, \cdots, \theta_k$, 由方程组

$$\begin{cases} \alpha_1 = f_1(\theta_1, \cdots, \theta_k) \\ \qquad\vdots \\ \alpha_k = f_k(\theta_1, \cdots, \theta_k) \end{cases}$$

反解得到

$$\begin{cases} \theta_1 = g_1(\alpha_1, \cdots, \alpha_k) \\ \qquad\vdots \\ \theta_k = g_k(\alpha_1, \cdots, \alpha_k) \end{cases}$$

将其中的总体矩用相应的样本矩代替，则我们可以得到参数$\theta_1, \cdots, \theta_k$的一个估计:

$$\begin{cases} \hat{\theta}_1 = g_1(a_1, \cdots, a_k) \\ \qquad\vdots \\ \hat{\theta}_k = g_k(a_1, \cdots, a_k) \end{cases}$$

若要估计参数$\theta_1, \cdots, \theta_k$的某函数$g(\theta_1, \cdots, \theta_k)$, 则用$g(\hat{\theta}_1, \cdots, \hat{\theta}_k)$去估计它.

这里我们用的都是原点矩α_k，当然也可以使用中心矩μ_k，或者两个都使用。在这种情况下，只需要把相应的总体矩换成样本矩。我们称这种估计方法为矩估计法，得到的估计量称为矩估计量。**矩估计方法应用的原则是：能用低阶矩处理的就不用高阶矩。**

矩估计法的优点是简单易行,并不需要事先知道总体是什么分布. 缺点是，当总体类型已知时，没有充分利用分布提供的信息. 一般场合下, 矩估计量不具有唯一性.

例 5.1.1. *投掷一枚硬币, 为了解正面出现的概率, 现独立重复的投掷n次, 用$X_1, \cdots, X_n$表示投掷结果. 显然此时总体X的分布为$B(1,p)$, p为感兴趣的量. 而$X_1, \cdots, X_n$为样本, 则求参数p的矩估计量。*

解: 由于$EX = p$，而样本均值$\bar{X}$收敛到总体均值EX, 因此p的一个矩估计量为$\hat{p} = \bar{X}$.

例 5.1.2. *为考察某种考试成绩分布情况, 使用正态分布$N(a, \sigma^2)$来作为总体X的分布. 现在从中随机调查n个人, 即样本为$X_1, \cdots, X_n$. 试求参数a, σ^2的矩估计量。*

解: 由于

$$EX = a, \quad Var(X) = \sigma^2$$

所以a, σ^2的一个矩估计量为

$$\hat{a} = \bar{X}, \quad \hat{\sigma}^2 = m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

我们知道$ES^2 = \sigma^2$，因此，σ^2的另一个矩估计量为$\hat{\sigma}^2 = S^2$.

§5.1.2 极大似然估计方法

极大似然方法到目前为止应用最广的的点估计方法. 这种方法是基于如下的看法:

定义 5.1.1. *设总体X 有概率函数$f(x;\theta) = f(x;\theta_1, \cdots, \theta_k)$，这里参数$\theta = (\theta_1, \cdots, \theta_k) \in \Theta$，而当固定$x$时把$f(x;\theta)$看成为$\theta$的函数，称为似然函数，常记为$L(x;\theta)$或$L(\theta)$.*

当固定参数θ时，$f(x;\theta)$可以看成是得到样本观察值x的可能性，这样，当把参数θ 看成变动时，也就得到"在不同的θ值下能观察到x的可能性大小, 即$L(x;\theta)$"；由于我们已经观察到了x，所以我们要寻求在哪一个θ的值下，使得能观察到x的可能性$L(x;\theta)$最大。这个θ的值即称为θ 极大似然估计值(看上去最有可能的)。我们先看一个例子:

例 5.1.3. *从鱼池里随机捕捞500条鱼, 做好记号后重新放入鱼池中, 待充分混合后再捕捞1000条鱼, 结果发现其中有72条带有记号. 试问鱼池中可能有多少条鱼.*

解: 先将问题一般化. 设池中有N条鱼, 其中r条做好记号. 随机捕捞s条, 发现x条有记号. 用上述信息来估计N.

用X表示捕捞的s条鱼中带记号鱼的数目, 则

$$P(X = x) = \frac{C_{N-r}^{s-x} C_r^x}{C_N^s}.$$

目前发现在捕捞的s条鱼中有记号的鱼x条, 要寻求N取何值时, 使得观察到这个事件$\{X = x\}$的可能性最大. 即x是固定的, N是变化的, 记$p(x;N) = P(X = x)$. 因为

$$g(N) := \frac{p(x;N)}{p(x;N-1)} = \frac{(N-s)(N-r)}{N(N-r-s+x)} = \frac{N^2 - N(s+r) + rs}{N^2 - N(r+s) + Nx},$$

当$rs > Nx$时, $g(N) > 1$; $rs < Nx$时, $g(N) < 1$. 所以$P(X = x)$在$N = \frac{rs}{x}$附近达到最大, 注意到N 只能取正整数, 故N的最可能的估计即极大似然估计为

$$\hat{N} = \left\lceil \frac{rs}{x} \right\rceil.$$

其中$\lceil \ \rceil$表示下取整, 即小于该值的最大整数. 将题目中的数字代入,

$$\hat{N} = \left\lceil \frac{500 \times 1000}{72} \right\rceil = 6944.$$

即鱼池中的总的鱼数为6694条.

现给出极大似然估计的一般性定义:

定义 5.1.2. *设$X = (X_1, \cdots, X_n)$为从具有概率函数f的总体中抽取的样本, θ为未知参数或者参数向量. $x = (x_1, \cdots, x_n)$为样本的观察值。若在给定x时, 值$\hat{\theta} = \hat{\theta}(x)$满足下式*

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(x; \theta)$$

则称$\hat{\theta}$为参数θ的极大似然估计值, 而$\hat{\theta}(X)$称为参数θ的极大似然估计量。若待估参数为θ的函数$g(\theta)$, 则$g(\theta)$的极大似然估计量为$g(\hat{\theta})$。

求极大似然估计值相当于求似然函数的最大值。在简单样本的情况下,

$$L(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

而把似然函数的对数$l(\theta) = \log L(\theta)$称为对数似然函数(这是由于在一些情况下，处理对数似然函数更方便)

当似然函数对变量θ单调时, 我们可以容易得到其最大值点. 反之当似然函数为非单调函数且对变量θ可微分时，我们可以求其驻点: 令

$$\frac{dl(\theta)}{d\theta} = 0 \quad (\text{或者}\frac{dL(\theta)}{d\theta} = 0)$$

当θ为多维时, 比如$\theta = (\theta_1, \cdots, \theta_k)$时令

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0 \quad (\text{或者}\frac{\partial L(\theta)}{\partial \theta_i} = 0)\ i = 1, \cdots, k$$

然后判断此驻点是否是最大值点。

例 5.1.4. 设$X_1, \cdots, X_n$为从总体$X \sim N(a, \sigma^2)$中抽取的样本，求参数a, σ^2的极大似然估计量。

解: 易得对数似然函数为

$$l(a, \sigma^2) = c - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2 - \frac{n}{2}log(\sigma^2)$$

其中c是与参数无关的常数. 令

$$\begin{cases} \frac{\partial l(a,\sigma^2)}{\partial a} = 0 \\ \frac{\partial l(a,\sigma^2)}{\partial \sigma^2} = 0 \end{cases}$$

得到

$$\begin{cases} a = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ \sigma^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - a)^2 \end{cases}$$

容易验证此驻点是唯一的最大值点，因此得到a, σ^2的极大似然估计量:

$$\begin{cases} \hat{a} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2. \end{cases}$$

有时函数f并不对$\theta_1, \cdots, \theta_k$可导, 甚至$f$本身也不连续, 这时求导就没法用, 必须回到原始定义.

例 5.1.5. 设总体X服从$[a, b]$上的均匀分布, $a < b$, 求参数a, b的极大似然估计.

解: 易得似然函数为

$$L(a, b) = \frac{1}{(b-a)^n}\prod_{j=1}^{n} I(a \le x_j \le b) = \frac{1}{(b-a)^n} I(a \le x_{(1)} \le x_{(n)} \le b).$$

于是对任何满足条件$a \le x_j \le b$的a, b都有

$$L(a, b) = \frac{1}{(b-a)^n} \le \frac{1}{(x_{(n)} - x_{(1)})^n},$$

即似然函数$L(a, b)$在$a = x_{(1)}, b = x_{(n)}$时取到最大值. 于是a, b的极大似然估计量为$\hat{a} = X_{(1)}, \hat{b} = X_{(n)}$.

例 5.1.6. 设$X_1,\cdots,X_n$为从具有如下形式密度的总体中抽取的样本:

$$f(x;a,b)=\begin{cases}\frac{1}{b}\exp\{-\frac{x-a}{b}\} & ,x>a\\ 0 & ,x\le a\end{cases}$$

求参数a,b的极大似然估计量.

解: 易得似然函数为

$$L(a,b)=\prod_{i=1}^{n}f(x_i;a,b)=\frac{1}{b^n}\exp\{-\frac{1}{b}\sum_{i=1}^{n}(x_i-a)\}I(x_{(1)}>a)$$

在固定b时，显然似然函数为a的单调增函数，因此$L(a)$的驻点为$\hat{a}=x_{(1)}$。再令$\frac{\partial L(a,b)}{\partial b}=0$，得到$b=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-x_{(1)})$，容易验证此解是最大值点。从而得到$a,b$的极大似然估计量:

$$\begin{cases}\hat{a}=X_{(1)}\\ \hat{b}=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-X_{(1)}).\end{cases}$$

例 5.1.7. 设$X_1,\cdots,X_n$为从如下分布中抽取的简单样本，求θ的极大似然估计.

$$f(x)=\frac{1}{x!(2-x)!}[\theta^x(1-\theta)^{2-x}+\theta^{2-x}(1-\theta)^x],\quad x=0,1,2;\quad \theta\in(0,\frac{1}{2})$$

解: 由题设知$f(x)$为离散型，其分布律为

X	0	1	2
P	$\frac{1}{2}[(1-\theta)^2+\theta^2]$	$2\theta(1-\theta)$	$\frac{1}{2}[(1-\theta)^2+\theta^2]$

若直接从此分布出发，则不能得到θ的极大似然估计的显式表达。为此，我们重新参数化，记$\eta=2\theta(1-\theta)$. 则由题设知$\eta<1/2$。则

X	0	1	2
P	$\frac{1}{2}(1-\eta)$	η	$\frac{1}{2}(1-\eta)$

再记$n_i=\#\{X_1,\cdots,X_n中等于i的个数\},\quad i=0,1,2,$ 则得到似然函数为

$$L(\eta)=(\frac{1}{2}(1-\eta))^{n_0}\eta^{n_1}(\frac{1}{2}(1-\eta))^{n_2}=(\frac{1}{2}(1-\eta))^{n-n_1}\eta^{n_1}$$

求解并注意η的上界即得到η的极大似然估计为

$$\hat{\eta}=\max\{\frac{n_1}{n},\frac{1}{2}\}$$

再由$\theta=\frac{1-\sqrt{1-2\eta}}{2}$得到$\theta$的极大似然估计为

$$\hat{\theta}=\frac{1-\sqrt{1-2\hat{\eta}}}{2}$$

§5.1.3 点估计的优良准则

我们看到对同一个参数，有多个不同的估计量，因此，评选不同估计量的优劣性是需要考虑的。

1. 无偏性

设$\hat{g}(X_1,\cdots,X_n)$为待估参数函数$g(\theta)$的一个估计量，若

$$E\hat{g}(X_1,\cdots,X_n)=g(\theta)$$

则称$\hat{g}(X_1,\cdots,X_n)$为$g(\theta)$的无偏估计量。无偏性是对一个估计量的最基本的要求, 其实际意义就是无系统误差. 因此在有多个估计量可供选择时，我们优先考虑无偏估计量。

很多时候我们得到的估计量是有偏, 例如正态总体的方差σ^2的极大似然估计量$\hat{\sigma}^2=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$是有偏的, $E\hat{\sigma}^2=\frac{n-1}{n}\sigma^2$. 若以$\frac{n}{n-1}$乘以$\hat{\sigma}^2$, 所得到的估计量就是无偏的. 这种方法称为修正.

若某一参数存在多个无偏估计时, 如何来选择使用哪个估计量? 人们又在无偏性的基础上增加了对方差的要求.

2. 有效性

设$\hat{g}_1(X_1,\cdots,X_n)$和$\hat{g}_2(X_1,\cdots,X_n)$为待估参数函数$g(\theta)$的两个不同的无偏估计量，若对任意的$\theta\in\Theta$,有

$$Var(\hat{g}_1(X_1,\cdots,X_n))\leq Var(\hat{g}_2(X_1,\cdots,X_n))$$

而且至少对某个$\theta_0\in\Theta$使得严格不等式成立。则称$\hat{g}_1$较$\hat{g}_2$有效。

3. 相合性

设总体分布依赖于参数$\theta_1,\cdots,\theta_k$, $g(\theta_1,\cdots,\theta_k)$是待估参数函数。设$X_1,\cdots,X_n$为自该总体中抽取的样本，$T(X_1,\cdots,X_n)$为$g(\theta_1,\cdots,\theta_k)$的一个估计量，如果对任意的$\epsilon>0$和$\theta_1,\cdots,\theta_k$的一切可能值都有

$$\lim_{n\to\infty}P_{\theta_1,\cdots,\theta_k}(|T(X_1,\cdots,X_n)-g(\theta_1,\cdots,\theta_k)|\geq\epsilon)=0$$

我们则称$T(X_1,\cdots,X_n)$为$g(\theta_1,\cdots,\theta_k)$的一个(弱)相合估计量。

相合性是对一个估计量的最基本的要求，如果一个估计量没有相合性，那么无论样本大小多大，我们也不能把未知参数估计到任意预定的精度。这种估计量显然是不可取的。

矩估计量是满足相合性的，极大似然估计量在很一般的条件下也是满足相合性的。

4. 渐近正态性

估计量是样本$X_1,\cdots,X_n$的函数，其确切的分布一般不是容易得到。但是，许多形式很复杂的统计量(未必是和)，当n很大时，其分布都渐近于正态分布，这个性质称为统计量的“渐近正态性”。

无偏性和有效性都是对固定的样本大小n而言的，这种性质称为估计量的“小样本性质”，而相合性和渐近正态性都是考虑在样本大小趋于无穷时的性质，这种性质称为“大样本性质”。

例 5.1.8. 设从总体

X	0	1	2	3
P	$\theta/2$	θ	$3\theta/2$	$1-3\theta$

抽取的一个简单样本$X_1,\cdots,X_{10}$的观察值为$(0,3,1,1,0,2,0,0,3,0)$，

(1) 求θ的矩估计量$\hat{\theta}_M$和极大似然估计量$\hat{\theta}_L$，并求出估计值。

(2) 上述估计量是否为无偏的？若不是，请作修正.

(3) 比较修正后的两个估计量，指出那个更有效.

由有效性的定义，我们自然会问在一切可能的无偏估计里，能否找到具有最小方差的无偏估计量？如果存在这样的估计量，我们称其为最小方差无偏估计量，详细地可以参考课本。

§5.2 区间估计

对于一个未知量,人们在测量和计算时，常不以得到近似值为满足，还需要估计误差，及要求知道近似值的精确程度(亦即所求真值所在的范围). 类似的，对于未知的参数θ，除了求出它的点估计$\hat{\theta}$外，我们还希望估计出一个范围，并希望知道这个范围包含参数θ真值得可信程度. 这样的范围通常以区间形式给出，同时还给出此区间包含真值得可信程度. 这种形式的估计称为区间估计.

比如一个人的年龄在18-25之间; 月支出在400-600元之间等. 区间估计的好处是把可能的误差用醒目的形式表示出来了. 比如你估计月花费支出是500，我们相信多少会有误差，但是误差有多大? 单从你提出的500这个数字还给不出什么信息，若你给出估计支

出是400-600之间, 则人们相信你在作出这估计时, 已把可能出现的误差考虑到了, 多少给人们以更大的信任感. 因此区间估计也是常用的一种估计方式.

现在最流行的一种区间估计理论是J. Neyman在上世纪30年代建立起来的. 他的理论的基本概念很简单, 为表达方便, 我们暂时假定总体分布只包含一个未知参数θ, 且要估计的就是θ本身. 如果总体分布中包含若干位置参数$\theta_1, \cdots, \theta_k$, 而要估计的是$g(\theta_1, \cdots, \theta_k)$, 则基本概念和方法并无不同. 这在后面的例子里可以看出.

§5.2.1 置信区间

Neyman建立起来的区间估计也叫置信区间, 字面上的意思是: 对该区间能包含未知参数θ可置信到何种程度. 假设$X_1, \cdots, X_n$是从该总体中抽取的样本, 所谓θ的区间估计,就是要寻求满足条件$\underline{\theta}(X_1, \cdots, X_n) < \bar{\theta}(X_1, \cdots, X_n)$的两个统计量$\underline{\theta}$和$\bar{\theta}$为端点的区间$[\underline{\theta}, \bar{\theta}]$. 一旦有个样本$X_1, \cdots, X_n$的值后,就把$\theta$估计在区间$[\underline{\theta}(X_1, \cdots, X_n), \bar{\theta}(X_1, \cdots, X_n)]$之内. 不难理解,这里有两个要求

- θ 以很大概率被包含在区间$[\underline{\theta}, \bar{\theta}]$内, 也就是说
$$P_\theta(\underline{\theta} \le \theta \le \bar{\theta}) = 1 - \alpha$$
  尽可能大, 即要求估计尽量可靠.

- 估计的精度要尽可能高，比如要求区间$[\underline{\theta}, \bar{\theta}]$要尽可能的短, 或者某种能体现这个要求的其他准则。

比如估计一个人的年龄, 如[30,35], 我们自然希望这个人的年龄有很大把握在这个区间之内, 并且希望这个区间不能太长. 如果估计是[10,90], 当然可靠了,但是精度太差, 用处不大.

但这两个要求是相互矛盾的，因此区间估计的原则是在已有的样本资源限制下，找出更好的估计方法以尽量提高可靠性和精度。Neyman 提出了广泛接受的准则：**先保证可靠性，在此前提下尽可能提高精度**。为此，引入如下定义:

定义 5.2.1. 设总体分布$F(x, \theta)$含有一个或多个未知的参数θ, $\theta \in \Theta$, 对给定的值$\alpha, (0 < \alpha < 1)$, 若由样本$X_1, \cdots, X_n$确定的两个统计量$\bar{\theta} = \bar{\theta}(X_1, \cdots, X_n)$ 和$\underline{\theta} = \underline{\theta}(X_1, \cdots, X_n)$, 满足
$$P_\theta(\underline{\theta} \le \theta \le \bar{\theta}) = 1 - \alpha \qquad \forall\ \theta \in \Theta$$
称$1 - \alpha$为置信系数或置信水平, 而称$[\underline{\theta}, \bar{\theta}]$为$\theta$的置信水平为$1 - \alpha$的置信区间。

区间估计就是在给定的置信水平之下，去寻找有优良精度的区间。

一般，我们首先寻求参数θ的一个估计(多数是基于其充分统计量构造的)，然后基于此估计量构造参数θ的置信区间，介绍如下:

1. 枢轴变量法 设待估参数为$g(\theta)$，

1. *找一个与待估参数$g(\theta)$有关的统计量T，一般是其一个良好的点估计(多数是通过极大似然估计构造);*

2. *设法找出T与$g(\theta)$的某一函数$S(T, g(\theta))$的分布，其分布F要与参数θ无关(S即为枢轴变量);*

3. *对任何常数$a < b$，不等式$a \leq S(T, g(\theta)) \leq b$要能表示成等价的形式$A \leq g(\theta) \leq B$，其中$A, B$只与$T, a, b$有关而与参数无关;*

4. *取分布F的上$\alpha/2$分位数$\omega_{\alpha/2}$和上$(1-\alpha/2)$分位数$\omega_{1-\alpha/2}$，有$F(\omega_{\alpha/2}) - F(\omega_{1-\alpha/2}) = 1-\alpha$. 因此*
$$P(\omega_{1-\alpha/2} \leq S(T, g(\theta)) \leq \omega_{\alpha/2}) = 1-\alpha$$

由3我们就可以得到所求的置信区间.

例 5.2.1. 设$X_1, \cdots, X_n$为从正态总体$N(\mu, \sigma^2)$中抽取得样本，求参数μ, σ^2的$1-\alpha$置信区间。

解：由于μ, σ^2的估计$\bar{X}, S^2$满足

$$T_1 = \sqrt{n}(\bar{X} - \mu)/S \sim t_{n-1}$$
$$T_2 = (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$$

所以T_1, T_2就是我们所要寻求的枢轴变量，从而易得参数μ, σ^2的$1-\alpha$置信区间分别为

$$\left[\bar{X} - \frac{1}{\sqrt{n}} S t_{n-1}(\alpha/2), \bar{X} + \frac{1}{\sqrt{n}} S t_{n-1}(\alpha/2)\right],$$

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right].$$

例 5.2.2. 设$X_1,\cdots,X_n$为从正态总体$N(\mu_1,\sigma_1^2)$中抽取得样本，$Y_1,\cdots,Y_m$为从正态总体$N(\mu_2,\sigma_2^2)$中抽取得样本，两组样本相互独立。求参数$\mu_1-\mu_2,\sigma_1^2/\sigma_2^2$的$1-\alpha$置信区间。

解: 方法完全类似于前面的例子，此处略.

2. 大样本法

大样本法就是利用中心极限定理，以建立枢轴变量。通过以下例子说明:

例 5.2.3. 某事件A在每次实验中发生的概率都是p，作n次独立的实验，以Y_n记A发生的次数。求p的$1-\alpha$置信区间。

解: 设n比较大，令$q=1-p$, 则由中心极限定理知，近似有$(Y_n-np)/\sqrt{npq}\sim N(0,1)$，从而$(Y_n-np)/\sqrt{npq}$可以作为枢轴变量。由

$$P(-u_{\alpha/2}\le (Y_n-np)/\sqrt{npq}\le u_{\alpha/2})\approx 1-\alpha \qquad (*)$$

可以等价表示成

$$P(A\le p\le B)\approx 1-\alpha$$

其中A,B为方程

$$(Y_n-np)/\sqrt{npq}=u_{\alpha/2}$$

的解, 即

$$A,B=\frac{n}{n+u_{\alpha/2}^2}\left[\hat{p}+\frac{u_{\alpha/2}^2}{2n}\pm u_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{u_{\alpha/2}^2}{4n^2}}\;\right]$$

A取负号，B取正号，$\hat{p}=Y_n/n$。

由于(*)式只是近似成立，故区间估计也只是近似成立，当n较大时才相去不远。详细的说明参见课本p203。我们还可以先假定方差是“已知”的，最后再将其估计，得到如下Wald 置信区间：

$$\hat{p}\pm u_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}.$$

§5.2.2 置信界

在实际中，有时我们只对参数θ的一端的界限感兴趣。比如果汁的最低含量, 有害物质的最高含量等等. 这就需要寻求参数θ的感兴趣的置信界限.

定义 5.2.2. 设总体分布$F(x,\theta)$含有一个未知的参数θ, $\theta \in \Theta$, 对给定的值$\alpha, (0 < \alpha < a)$, 若由样本$X_1, \cdots, X_n$确定的两个统计量$\bar{\theta} = \bar{\theta}(X_1, \cdots, X_n)$ 和$\underline{\theta} = \underline{\theta}(X_1, \cdots, X_n)$,

*1.*若

$$P_\theta(\theta \leq \bar{\theta}) \geqq 1 - \alpha \qquad \forall\ \theta \in \Theta$$

则称$\bar{\theta}$为θ的一个置信系数为$1-\alpha$的置信上界。

*2.*若

$$P_\theta(\theta \geq \underline{\theta}) \geqq 1 - \alpha \qquad \forall\ \theta \in \Theta$$

则称$\underline{\theta}$为θ的一个置信水平为$1-\alpha$的置信下界。

而$(-\infty, \bar{\theta}]$ 和$[\underline{\theta}, +\infty)$都称为是单边的置信区间。

寻求置信上、下界的方法和寻求置信区间的方法完全类似。

§5.2.3 确定样本大小

在以区间长度为精度准则下, 置信区间越窄就越好，为什么呢? 作为一个一般的原则，我们已经知道更多的测量可以得到更精确的推断。有时候，对精度是有要求的，甚至于是在测量之前就提出此要求，因此相应的样本大小就要事先确定下来。我们以如下的例子说明如何确定样本大小，一般的方法类似。

例 5.2.4. 假设某种成分的含量服从正态分布$N(\mu, \sigma^2)$, σ^2已知。要求平均含量μ的$(1-\alpha)$置信区间的长度不能长于ω。试确定测量样本大小。

解: 由于σ^2已知，我们已经知道可以根据$\bar{X} \sim N(\mu, \sigma^2/n)$来构造$\mu$的$95\%$ 置信区间。因此易知区间长度为$2u_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. 从而由

$$2u_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \omega$$

得到

$$n \geq \left(\frac{2u_{\alpha/2}\sigma}{\omega}\right)^2.$$

比如当$\sigma = 0.1, \omega = 0.05, \alpha = 0.05$, 可以得到$n \geq \left(\frac{2\times 1.96 \times 0.1}{0.05}\right)^2 = 61.4656$. 即为达到要求至少需要测量62次。

第六章 假设检验

教学目的:

1) 理解假设检验的一些基本概念: 零假设、对立假设、两类错误、拒绝域、显著性水平、功效.

2) 学会将实际问题转化成假设检验问题来处理.

3) 一样本和两样本正态总体均值和方差的假设检验.

4) 0-1 分布参数的假设检验.

5) 拟合优度检验、列联表的独立性和齐一性检验.

§6.1 基本概念和问题的提法

§6.1.1 零假设, 对立假设, 两类错误, 拒绝域, 显著性水平, 功效

在参数估计问题中, 常常在抽样前先对未知总体作一些假定. 例如假定总体X 服从正态分布, 假定某个正态总体的方差为一个已知值等等. 在数理统计中, 关于总体分布的概率性质的假定称为**(统计) 假设**. 抽样前所作出的假设是否与实际符合, 可以用样本所提供的信息来检查, 检查的方法与过程称为**(统计) 检验**. **假设检验问题**就是研究如何根据抽样后获得的样本来检验抽样前所作出的假设. 首先, 由一个例子引出一些基本概念.

例 6.1.1. *某饮料厂在自动流水线上罐装饮料. 在正常生产情况下, 每瓶饮料的容量(单位: 毫升) X 服从正态分布$N(500, 10^2)$ (由以往的经验得知). 经过一段时间之后, 有人觉得每瓶饮料的平均容量减小到490, 于是抽取了9 瓶样品, 称得它们的平均值为$\bar{x} = 492$ 毫升. 试问此断言是否正确? 即问平均每瓶饮料的容量仍是500 毫升还是变成490 毫升? 假定标准差10 毫升不变.*

在这个问题中, 设经过一段时间后罐装饮料容量X 的平均值为μ, 则由题意可设$X \sim N(\mu, 10^2)$. 记$x_1, \cdots, x_9$ 为取自这个正态总体X 的一组样本观测值, 则$\bar{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = 492$. 我们需要在“饮料平均容量为500 毫升”与“饮料平均容量为490 毫升”之间作判断,

即在“$\mu = 500$”和“$\mu = 490$”之间作判断. 数理统计中, 把它们看成两个假设. 习惯上, 称前者为**原假设或零假设**, 记作H_0; 后者称为**备择假设或对立假设**, 记作H_1 或H_a. 所谓检验

$$H_0 : \mu = 500 \leftrightarrow H_1 : \mu = 490.$$

就是要根据样本判断究竟是“H_0成立”还是“H_1成立”. 断言“H_0成立”称为**接受**H_0; 断言“H_1成立”称为**拒绝**H_0.

下面讨论如何检验上述假设, 即给定一个接受或者拒绝零假设的准则. 设从总体中抽取一个样本$X_1, \cdots, X_n$, 我们可以用极大似然估计$T = \bar{X}$ (称之为**检验统计量**) 来估计μ. 由于该估计值接近μ (尤其是当样本量较大时), 故当T 的绝对值小的时候有利于H_1 而不利于H_0, 此时应该拒绝H_0. 我们可以事先取定一个常数τ, 称之为**临界值**, 当T 的取值小于该临界值时拒绝H_0, 即样本满足

$$W = \{\bar{X} < \tau\}$$

中时拒绝H_0, 称W 为**拒绝域**. 即样本的取值落在拒绝域中, 就拒绝H_0, 否则不能拒绝之. 一个拒绝域就对应于一个检验方法. 现在的问题是τ 应该取多大? 这涉及到**两类错误**.

事实 决策	H_0 成立	H_1 成立
接受H_0	不犯错	第II 类错误
拒绝H_0	第I 类错误	不犯错

称“实际上H_0 成立但是它被拒绝”这个错误为**第I 类错误(弃真)** , 而“实际上H_0 不成立但是它被接受”这样一类错误为**第II 类错误(存伪)**. 由于我们的方法是基于观测数据, 而观测数据是带有随机误差的, 故难免在做出决策的时候犯错, 我们能做的是控制犯错的概率. 一个理想的检验应该使这两类错误的概率都小, 但是在实际问题中不可能使这两类错误一致地小: 要让犯第I 类错误的概率小, 应该让τ 小, 而要让犯第II 类错误的概率小, 则τ 不能太小. 解决这个矛盾的一个方法是**在控制I类错误的基础上, 尽量少犯第II 类错误** (在下一小节中我们讨论如何设定假设时会提到, 应该将受保护对象设为零假设, 故犯第I 类错误的严重性更大, 因此必须尽量避免犯第I 类错误). 具体地, 选定一个小的常数α, 取τ 使得犯第I 类错误的概率, 即T 小于τ 的概率小于α. 称α 为**显著性水平**. 理想情况下, τ 取得恰好满足$P_{H_0}(T < \tau) = \alpha$. 为控制犯第I 类错误的发生, 通常将$\alpha$ 取为0.1,

0.05, 0.01 等较小的数, 具体取值视实际需要而定, 有时候要求α 很小, 比如在涉及到数十万个基因标记的基因关联分析中, 单个位点检验的α 一般是10^{-7} 这样的量级.

现在将问题一般化. 设有假设检验问题

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1. \tag{6.1.1}$$

其中H_0 为**零假设**或**原假设** 而H_1 为**对立假设**或**备择假设**. 构造一个适当的**检验统计量** $T = T(X_1, \cdots, X_n)$, 其中$X_1, \cdots, X_n$ 为从总体中抽得的一个样本. 根据对立假设的形状构造一个检验的**拒绝域** $W = \{T(X_1, \cdots, X_n) \in A\}$, 其中$A$ 为一个集合, 通常是一个区间. 比如拒绝域可取为$\{T(X_1, \cdots, X_n) > \tau\}$, 则称$\tau$ 为**临界值**. 如果零假设成立但拒绝了零假设, 则称犯了**第I 类错误**, 如果对立假设成立但接受零假设, 则称犯了**第II 类错误**. 如对任意的$\theta \in \Theta_0$, 犯第I 类错误的概率$P_\theta(T(X_1, \cdots, X_n) \in A)$ 小于或等于某个正的常数α), 则称α 为**显著性水平**. 显然显著性水平不是唯一的, 事实上, 如果α 是一个显著性水平, 则任意大于α 的数都是显著性水平. 实际中通常采用显著性水平最小的那一个. 一个**检验** 对应于一个拒绝域, 称$\beta(\theta) = P_\theta$ (H_0 被拒绝) 为检验的**功效函数**. 如果检验的显著性水平为α, 则当$\theta \in \Theta_0$ 时, $\beta(\theta) \le \alpha$. 而当$\theta \in \Theta_1$ 时, 我们希望功效值越大越好(这样犯第II 类错误的概率$1 - \beta(\theta)$ 就越小), 所以功效可以作为评价一个检验优劣的准则.

§6.1.2 假设检验问题的提法

在有时候需要自己判断如何提假设检验问题. 在建立假设检验问题时有两个原则。

原则一: 将受保护的对象置为零假设. 如我国按照以前的司法制度, 公安机关抓到嫌疑犯后, 很多情况下要犯人自己证明无罪(有罪推断) , 这对嫌疑犯很不利, 从而容易导致冤案. 现在的司法制度则总假定嫌疑犯是无罪的, 要司法部门证明其有罪(无罪推断), 这样做大大地有利于保护公民的利益, 如果要将真正的嫌疑犯绳之以法, 则司法部门必须有充分的证据, 这样做可以有效保护公民的权益, 对司法部门要求也变高了. 又比如药厂生产出一种新药, 在上市前要通过食品与药品监管局的检验. 显然使用药品的病人是应该受保护的对象, 这时应该设定一个有利于病人的命题作为零假设, 这个命题就是“新药不比安慰剂效果好”, 以尽量避免病人用无效甚至有副作用的新药. 当然, 对立假设就是“新药比安慰剂效果好’. 将检验的显著性水平α 设定得较小, 以保证零假设不被轻易推翻. 在实际问题中, 如果根据某个合理的检验方法发现零假设被推翻, 则有充分的理由认为零假设不成立而对立假设成立, 这是因为万一零假设成立而被误据的概率不会超过α; 另一方面, 如果发现零假设未被拒绝, 并不表明有充分理由接受零假设, 而

是因为零假设被保护得较严密以至于未被拒绝.

原则二: 如果你希望"证明"某个命题, 就取相反结论或者其中一部分作为零假设(类似于反证法). 这种提法往往是在两个假设命题中不太清楚哪个应受保护, 此时可以借用司法制度里的"谁主张, 谁举证", 即若想用统计方法向人"证明" 一个命题, 则将那个命题置为对立假设. 注意这里的证明不是数学上的严格证明, 而是允许犯错的一种统计推断方法. 用统计方法证明一个命题不是一件容易的事情, 所以如果没有足够把握, 人们应该避免用统计方法去证明一个命题.

上述两原则是统一的: 一般不应该让受保护对象去证明一个命题.

§6.1.3 检验统计量的选取及假设检验的步骤

通过解答例6.1.1来说明假设检验的步骤.

例 6.1.2. *(例6.1.1续) 能否在显著性水平0.05 下认为饮料的平均容量确实减少到490 毫升?*

解: 基于统计量$\bar{X}$, 我们采用"标准化"过的检验统计量(减均值再除以标准差)

$$T_1 = \frac{\sqrt{n}(\bar{X} - 500)}{10}$$

以使该统计量服从标准正态分布, 检验的拒绝域仍取形如$\{T_1 < \tau_1\}$, 我们控制犯第I 类错误的概率等于α, 即

$$P(T_1 < \tau_1 | \theta = 500) = \alpha.$$

由于$\theta = 500$ 时T_1 服从标准正态分布, 易知上面关于τ_1 的方程的解为$\tau_1 = -u_\alpha$, 其中u_c 等于标准正态分布的上c 分位数, 即检验的拒绝域为

$$\{T_1 < -u_\alpha\}.$$

现在取显著性水平为0.05, 则临界值$u_{0.05} \approx 1.645$. 另一方面, 样本均值$\bar{x} = 492$, 样本量$n = 9$, 故检验统计量T_1 的观测值等于-2.4, 小于临界值1.645, 即样本落在拒绝域中, 从而可以在显著性水平0.05 下拒绝零假设, 认为饮料的平均容量确实减少为490 毫升.

下面列举几种常见的假设检验问题:

(1) $H_0: \ \theta = \theta_0 \leftrightarrow H_1 : \theta = \theta_1$;

(2) $H_0: \ \theta = \theta_0 \leftrightarrow H_1 : \ \theta \neq \theta_0$;

(3) $H_0:\ \theta=\theta_0 \leftrightarrow H_1:\ \theta>\theta_0$或者$H_0:\ \theta\le\theta_0 \leftrightarrow H_1:\ \theta>\theta_0$

(4) $H_0:\ \theta=\theta_0 \leftrightarrow H_1:\ \theta<\theta_0$或者$H_0:\ \theta\ge\theta_0 \leftrightarrow H_1:\ \theta<\theta_0$

称(1) 为**简单假设**, (2)为**双侧假设**因为对立假设是双侧的, (3) 和(4) 为**单侧假设**因为对立假设是单侧的. 这里强调对立假设的原因是检验方法(对应于一个拒绝域) 只跟对立假设有关.

下面我们给出检验上述假设的一般步骤, 它的基本思想是: 一个好的点估计应该是一个优良检验的的主要依据, 设定显著性水平为α.

第1 步: 求出未知参数θ 的一个较优的点估计$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$, 如极大似然估计.

第2 步: 以$\hat{\theta}$ 为基础, 寻找一个检验统计量

$$T=t(X_1,\cdots,X_n)$$

且使得当$\theta=\theta_0$ 时, T 的分布已知(如$N(0,1), t_n, F_{m,n}$) , 从而容易通过查表或计算得到这个分布的分位数, 用以作为检验的临界值.

第3 步: 以检验统计量T 为基础, 根据对立假设H_1 的实际意义, 寻找适当形状的拒绝域, 它是关于T 的一个或两个不等式), 其中包含一个或两个临界值.

第4 步: 当零假设成立时, 犯第I 类错误的概率小于或等于给定的显著性水平α, 这给出一个关于临界值的方程, 解出临界值, 它(们) 等于T 的分位数, 这样即确定了检验的拒绝域.

第5 步: 如果给出样本观测值, 则可算出检验统计量的样本观测值, 如落在拒绝域中则可拒绝零假设, 否则不能.

§6.2 重要参数检验

本节介绍最基本的假设检验问题: 一样本和两样本正态总体的有关均值和方差的检验, 简单的大样本检验(0-1 分布参数的假设检验).

§6.2.1 一样本正态总体均值和方差的检验

现实中经常碰到诸如此类的问题: 假设用于某用途的合格铁钉要求长度为10 厘米, 现有经销商从生产厂家订购了一批这样的铁钉, 为了检验该批检验产品是否合格, 可以

从中抽取一小部分进行测量检验, 通常铁钉的长度服从一个正态分布, 这类问题属于一样本正态总体的假设检验问题.

一般地, 设总体$X \sim N(\mu, \sigma^2), -\infty < \mu < \infty, \sigma^2 > 0$; $X_1, \cdots, X_n$ 是取自总体X 的一个样本. 取显著性水平为α.

(1) 方差已知时均值的检验

先考虑双侧假设, 即要检验

$$H_0 : \mu = \mu_0 \leftrightarrow H_1 : \mu \neq \mu_0.$$

由于μ 的极大似然估计为$\bar{X}$, 取"标准化"后的检验统计量

$$U = u(X_1, \cdots, X_n) = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}$$

注意到当H_0 成立时, $U \sim N(0, 1)$, $|U|$ 应该较小, 反之当$|U|$ 的观测值$u(x_1, \cdots, x_n)$ 较大时, 不利于零假设H_0 应该拒绝之. 所以选拒绝域形如

$$\{|U| > \tau\}.$$

要求显著性水平为α, 即

$$P_{H_0}(|U| > \tau) = \alpha,$$

解得$\tau = u_{\alpha/2}$. 于是检验的拒绝域为

$$\{|U| > u_{\alpha/2}\}.$$

即当观测值$(x_1, \cdots, x_n)$ 满足不等式

$$\sqrt{n}\frac{|\bar{x} - \mu_0|}{\sigma} > u_{\alpha/2}$$

时拒绝H_0.

类似地, 检验单侧假设

$$H_0 :\ \mu = \mu_0 \leftrightarrow H_1 :\ \mu > \mu_0 \quad \text{或者} \quad H_0 :\ \mu \leq \mu_0 \leftrightarrow H_1 :\ \mu > \mu_0$$

仍然用统计量U, 由于U 大时不利于H_0, 取拒绝域为

$$\{U > u_\alpha\}.$$

而检验另一个单侧假设

$$H_0:\ \mu=\mu_0 \leftrightarrow H_1:\ \mu<\mu_0 \quad \text{或者} \quad H_0:\ \mu\le\mu_0 \leftrightarrow H_1:\ \mu<\mu_0$$

的拒绝域为

$$\{U<-u_\alpha\}.$$

虽然我们取的临界值只考虑使检验在$\mu=\mu_0$ 处的犯I 类错误的概率为α, 从检验的拒绝域的形状上可直接看出来在零假设下$\mu\le\mu_0$ (或$\mu\ge\mu_0$) 时犯第I 类错误的概率恒小于或等于α.

以上三个检验统称为u 检验.

例 6.2.1. *随机地从一批铁钉中抽取16 枚, 测得它们的长度(单位: 厘米) 如下:*

2.942371 2.988662 3.106234 3.109316 3.118427 3.132254 3.140042 3.170188

2.902562 3.128003 3.146441 2.978240 3.103600 3.003394 3.044384 2.849916

已知铁钉长度服从标准差为0.1 的正态分布, 在显著性水平$\alpha=0.01$ 下, 能否认为这批铁钉的平均长度为3 厘米? 如显著性水平为$\alpha=0.05$ 呢?

解: 这是方差已知时关于均值μ 的假设检验问题,

$$H_0:\mu=3 \leftrightarrow H_1:\mu\ne 3.$$

取检验统计量为$U=\sqrt{n}(\bar{X}-3)/0.1$, 检验的拒绝域为$|U|>u_{\alpha/2}$. 由样本算得检验统计量的值为$u\approx 2.16$, 如显著性水平为0.01, 则临界值为$u_{0.005}\approx 2.58$, 跟检验统计量的值比较发现不能拒绝零假设, 即不能推翻铁钉平均长度为3 厘米的假设; 而如果显著性水平为0.05时, 临界值为$u_{0.025}=1.96$, 此时可以拒绝零假设, 认为铁钉平均长度不等于3 厘米. 这个例子说明结论可能跟显著性水平的选择有关: 显著性水平越小, 零假设被保护得越好从而更不容易被拒绝.

(2) 方差未知时均值的检验

考虑检验

$$H_0:\mu=\mu_0 \leftrightarrow \mu\ne\mu_0,$$

由于方差未知, 可以在将$\bar{X}$ 标准化的过程中用样本方差S^2 代替总体方差σ^2, 得检验统计量

$$T=\sqrt{n}\frac{\bar{X}-\mu_0}{S}.$$

由于在H_0 下, $T \sim t_{n-1}$, 于是拒绝域取成

$$\{|T| > t_{n-1}(\alpha/2)\}.$$

此检验称为t 检验.

类似地可以得到另外两个单侧假设的检验拒绝域, 列于表6.2.1中.

例 6.2.2. *(例6.2.1续) 设方差未知, 则在水平0.01 和0.05 下能否认为铁钉平均长度为3 厘米?*

解: 这是方差未知时关于均值μ 的假设检验问题,

$$H_0 : \mu = 3 \leftrightarrow H_1 : \mu \neq 3$$

取检验统计量为$T = \sqrt{n}(\bar{X} - 3)/S$, 检验的拒绝域为$|T| > t_{n-1}(\alpha/2)$. 由样本算得检验统计量的值约为2.21, 与显著性水平0.01 对应临界值$t_{15}(0.005) \approx 2.95$ 比较, 不能拒绝零假设, 而与显著性水平0.05 对应临界值$t_{15}(0.025) \approx 2.13$ 比较, 可以拒绝零假设, 即在显著性水平0.01 下不能拒绝铁钉平均长度为3 厘米的假定, 但在显著性水平0.05 下可以认为铁钉平均长度不等于3 厘米, 此结论与方差已知情形一致.

(3) 方差的检验

考虑假设检验问题

$$H_0 : \sigma^2 = \sigma_0^2 \leftrightarrow H_1 : \sigma^2 \neq \sigma_0^2.$$

对均值已知的情形, 由σ^2 的极大似然估计

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

可以构造检验统计量

$$\chi^2 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{n\hat{\sigma}^2}{\sigma_0^2}.$$

在H_0 下, $\chi^2 \sim \chi_n^2$, χ^2 的平均值为n, 而在H_1 下, $\chi^2 = \frac{\sigma^2}{\sigma_0^2}\frac{n\hat{\sigma}^2}{\sigma^2}$ 的均值为$\frac{\sigma^2}{\sigma_0^2}n \neq n$, 因此当$\chi^2$ 的值过于偏离n 时应该拒绝H_0, 于是拒绝域取成

$$\left\{\chi^2 < \chi_n^2(1-\alpha/2) \text{ 或者 } \chi^2 > \chi_n^2(\alpha/2)\right\}.$$

对均值未知的情形, 构造检验统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2},$$

其中S^2 为样本方差. 在H_0 下, $\chi^2 \sim \chi^2_{n-1}$, 拒绝域取成

$$\left\{\chi^2 < \chi^2_{n-1}(1-\alpha/2) \text{ 或者 } \chi^2 > \chi^2_{n-1}(\alpha/2)\right\}.$$

对于单侧假设, 可以类似得到检验的拒绝域, 参看表6.2.1.

上述检验称为χ^2 检验.

例 6.2.3. *(例6.2.1续) 在水平0.1 下能否认为铁钉的标准差大于0.1 厘米?*

解: 这是均值未知时关于方差σ^2 的假设检验问题,

$$H_0 : \sigma^2 \le 0.1^2 \leftrightarrow H_1 : \sigma^2 > 0.1^2.$$

取检验统计量为$\chi^2 = \frac{(n-1)S^2}{0.1^2}$, 检验的拒绝域为$\{\chi^2 > \chi^2_{n-1}(\alpha)\}$. 由样本算得检验统计量的值$\chi^2 \approx 14.32$, 与显著性水平0.2 对应临界值$\chi^2_{15}(0.1) \approx 22.31$ 比较, 不能拒绝零假设, 即在显著性水平0.1 下可以认为铁钉的标准差小于0.1.

表6.2.1 总结了有关一样本正态总体的假设检验.

§6.2.2 两样本正态总体的情形

为了检验某肥料是否能显著提高玉米产量, 可以设计一个随机试验: 选择两块条件一样的试验区, 把两试验区各分成若干小块, 一个试验区的各小块施肥, 另一个试验区的各小块不施肥, 最后统计收成, 可以采用如下的检验方法来检验玉米产量差别, 从而知道肥料是否有效.

设总体$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $-\infty < \mu_1, \mu_2 < \infty, \sigma_1^2, \sigma_2^2 > 0$; $X_1, \cdots, X_n$ 是从总体X 中抽取的一个样本, $Y_1, \cdots, Y_n$ 是从总体Y 中抽取的一个样本. 设来自不同总体的样本相互独立. 下面设考虑有关均值差$\mu_1 - \mu_2$ 和方差比σ_1^2/σ_2^2 的检验. 取显著性水平为α. 举例说明.

例 6.2.4. *甲乙两个农业试验区种植玉米, 除了甲区施磷肥外, 其他试验条件都相同, 把两个试验区分别均分成10 个和9 个小区统计产量(单位: 千克), 得数据如下*

甲区 62 57 65 60 63 58 57 60 60 58

表 6.2.1 一样本正态总体$N(\mu,\sigma^2)$.

检验对象	检验统计量	分布	拒绝域†
μ (σ^2已知)	$U=\sqrt{n}(\bar{X}-\mu_0)/\sigma$	$N(0,1)$	$\begin{cases} \lvert U\rvert>u_{\alpha/2} \\ U>u_\alpha \\ U<-u_\alpha \end{cases}$
μ (σ^2 未知)	$T=\sqrt{n}(\bar{X}-\mu_0)/S$	t_{n-1}	$\begin{cases} \lvert T\rvert>t_{n-1}(\alpha/2) \\ T>t_{n-1}(\alpha) \\ T<-t_{n-1}(\alpha) \end{cases}$
σ^2 (μ 已知)	$\chi^2=\frac{1}{\sigma_0^2}\sum_{i=1}^n(X_i-\mu)^2$	χ_n^2	$\begin{cases} \chi^2>\chi_n^2(\alpha/2)\text{或者}\chi^2<\chi_n^2(1-\alpha/2) \\ \chi^2>\chi_n^2(\alpha) \\ \chi^2<\chi_n^2(1-\alpha) \end{cases}$
σ^2 (μ 未知)	$\chi^2=\frac{1}{\sigma_0^2}\sum_{i=1}^n(X_i-\bar{X})^2$	χ_{n-1}^2	$\begin{cases} \chi^2>\chi_{n-1}^2(\alpha/2)\text{或者}\chi^2<\chi_{n-1}^2(1-\alpha/2) \\ \chi^2>\chi_{n-1}^2(\alpha) \\ \chi^2<\chi_{n-1}^2(1-\alpha) \end{cases}$

†有关均值的检验: 对立假设分别为$\mu\neq\mu_0$, $\mu>\mu_0$ 和$\mu<\mu_0$. 有关方差的检验: 对立假设分别为$\sigma^2\neq\sigma_0^2$, $\sigma^2>\sigma_0^2$ 和$\sigma^2<\sigma_0^2$.

乙区 *50 59 56 57 58 57 56 55 57*

假定甲乙两区中每小块的玉米产量分别服从$N(\mu_1,\sigma^2)$, $N(\mu_2,\sigma^2)$, 其中μ_1,μ_2,σ^2 未知. 试问在显著性水平$\alpha=0.1$ 下磷肥对玉米的产量是否有效?

解: 磷肥对玉米产量有效果等价于$\mu_1>\mu_2$, 故将其设为对立假设, 假设检验问题是

$$H_0:\mu_1\leq\mu_2=0\leftrightarrow H_1:\mu_1>\mu_2.$$

构造基于$\mu_1-\mu_2$ 的极大似然估计$\bar{X}-\bar{Y}$ 的检验统计量

$$T=\frac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}.$$

当H_0 成立时, $T\sim t_{m+n-2}$, 于是拒绝域为

$$\{T>t_{m+n-2}(\alpha)\}.$$

由所得数据算得检验统计量T 的观测值为

$$t=\frac{\bar{x}-\bar{y}}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}=3.23.$$

由$\alpha = 0.1$ 得临界值为$t_{m+n-2}(\alpha/2) = t_{17}(0.1) \approx 1.33 < 3.23$, 因此拒绝$H_0$, 即可以在显著性水平0.1 下认为磷肥对玉米的产量有显著性影响.

例 6.2.5. *在例6.2.4中假定了两个正态总体的方差是相等的, 即$\sigma_1^2 = \sigma_2^2 = \sigma^2$. 现在我们根据样本来检验这个方差齐性的假设, 即要检验*

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \leftrightarrow H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1.$$

解: 因为σ_1^2 和σ_2^2 的极大似然估计分别是

$$\hat{\sigma}_1^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2, \hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

在$\theta = \sigma_1^2/\sigma_2^2$ 的极大似然估计$\hat{\theta} = \hat{\sigma}_1^2/\hat{\sigma}_2^2$ 的基础上可以构造检验统计量

$$F = \frac{S_1^2}{S_2^2} = \frac{(m-1)\hat{\sigma}_1^2/m}{(n-1)\hat{\sigma}_2^2/n}.$$

注意到F 中的分子和分母分别是X 和Y 的样本方差. 当零假设成立时, $F \sim F_{m-1,n-1}$. 于是拒绝域为

$$\{F < F_{m-1,n-1}(\alpha/2) \quad \text{或} \quad F > F_{m-1,n-1}(1-\alpha/2)\}.$$

由数据算得检验统计量F 的观测值$f = 1.19$, 如果取显著性水平$\alpha = 0.2$, 那么临界值为$F_{9,8}(0.1) = 2.44$, $F_{9,8}(0.9) = 1/F_{8,9}(0.1) = 0.41$ (如果$X \sim F_{m,n}$, 则$1/X \sim F_{n,m}$). 易见$0.41 < 1.19 < 2.44$, 因此不能拒绝H_0, 即在显著性水平0.2 下可以认为上例中所作的方差齐性假定是合理的.

表6.2.2 总结了两样本正态总体的双侧假设检验.

§6.2.3 成对数据

在上述两样本正态总体的假设检验中, 要求两个样本是独立的, 但是没有要求样本量相等. 有一类数据叫做成对数据$\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$, 比如一个病人在用药前后测得的指标分别为$X$ 和Y, 则X 与Y 总是一起出现的, 且由于它们是同一个体的指标, 故具有很大的相关性而绝对不是独立的, 这与两样本正态总体有本质区别. 另外, 两样本检验问题要求样本$X_1, \cdots, X_m$ 是同分布的($Y_1, \cdots, Y_n$ 亦然), 而成对数据则无此要求, 而要求$X_1 - Y_1, \cdots, X_n - Y_n$ 是同分布. 比如病人可以是来自两个不同性别、种族、年龄层的人. 要检验用药前后的指标有无显著差别, 可以构造一个新的总体$Z = Y - X$ 及样本$Z_1 = X_1 - Y_1, \cdots, Z_n = X_n - Y_n$, 相应的假设检验是一样本的! 在实际问题中, 如果发现有两个样本且其样本量是相等的, 则要检查独立性和同分布性, 否则可能是成对数据.

表 6.2.2 两样本正态总体的假设检验

检验对象	检验统计量	分布	拒绝域†
均值(方差已知)	$U=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$	$N(0,1)$	$\begin{cases}\lvert U\rvert>u(\alpha/2)\\ U>u(\alpha)\\ U<-u(\alpha)\end{cases}$
均值(方差未知)‡	$T=\frac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}$	t_{m+n-2}	$\begin{cases}\lvert T\rvert>t_{m+n-2}(\alpha/2)\\ T>t_{m+n-2}(\alpha)\\ T<-t_{m+n-2}(\alpha)\end{cases}$
方差(均值已知)	$F=\frac{\sum_{i=1}^m(X_i-\mu_1)^2/m}{\sum_{i=1}^n(X_i-\mu_2)^2/n}$	$F_{m,n}$	$\begin{cases}F>F_{m,n}(\alpha/2)\text{或}F<\frac{1}{F_{n,m}(\alpha/2)}\\ F>F_{m,n}(\alpha)\\ F<\frac{1}{F_{n,m}(\alpha)}\end{cases}$
方差(均值未知)	$F=\frac{S_1^2}{S_2^2}$	$F_{m-1,n-1}$	$\begin{cases}F>F_{m-1,n-1}(\alpha/2)\text{或}F<\frac{1}{F_{n-1,m-1}(\alpha/2)}\\ F>F_{m-1,n-1}(\alpha)\\ F<\frac{1}{F_{n-1,m-1}(\alpha)}\end{cases}$

†有关均值的检验: 对立假设分别为$\mu_1\neq\mu_2$, $\mu_1>\mu_2$ 和$\mu_1<\mu_2$. 有关方差的检验: 对立假设分别为$\sigma_1^2\neq\sigma_2^2$, $\sigma_1^2>\sigma_2^2$ 和$\sigma_1^2<\sigma_2^2$.
‡假定方差相等

§6.2.4 0-1 分布中未知参数p 的假设检验

产品验收时, 需要检验不合格率是否小于某给定的一个数.

设$(X_1,\cdots,X_n)$ 是取自总体X 的一个样本, 该总体服从0-1 分布, 取1 的概率为p. 常见的假设有三种:

(1) $H_0:p=p_0\leftrightarrow H_1:p\neq p_0$;

(2) $H_0:p=p_0\leftrightarrow H_1:p>p_0$ 或$H_0:p\leq p_0\leftrightarrow H_1:p>p_0$;

(3) $H_0:p=p_0\leftrightarrow H_1:p<p_0$ 或$H_0:p\geq p_0\leftrightarrow H_1:p<p_0$.

假定样本量n 较大, 取显著性水平为α. 由于p 的极大似然估计为$\bar{X}$, 取"标准化"过的检验统计量

$$T=\sqrt{n}\frac{\bar{X}-p_0}{\sqrt{p_0(1-p_0)}},$$

其中p_0 和$p_0(1-p_0)/n$ 分别为$\bar{X}$ 在零假设$p=p_0$ 下的期望和方差, 从而当H_0 成立时, 由中心极限定理近似地有$T\sim N(0,1)$. 于是上述三种检验的拒绝域分别为

$$\{|T|>u_{\alpha/2},\quad \{T>u_\alpha\}\quad \text{和}\quad \{T<-u_\alpha\}$$

例 6.2.6. *某厂产品不合格率通常为0.5. 厂方希望知道原料产地的改变是否对产品的质量发生显著的影响. 现在随机地从原料产地改变后的产品中抽取了80 个样品进行检验, 发现有5 个是不合格品. 试问, 在显著性水平0.1 下, 厂方由此可以得出什么结论?*

解: 总体$X \sim B(1,p)$, 其中p 未知. 在显著性水平$\alpha = 0.1$ 下, 产品质量无变化等价于$p = 0.05$, 故我们要检验

$$H_0 : p = 0.05 \leftrightarrow H_1 : p \neq 0.05.$$

由于$\bar{x} = 5/80 = 0.0625$, 因此检验统计量T 的观测值

$$t = \sqrt{n}\frac{\bar{x} - p_0}{\sqrt{p_0(1-p_0)}} = 0.513.$$

由$\alpha = 0.10$ 得临界值$u_{0.05} = 1.645$. 易见, $|t| < 1.645$, 因此不能拒绝H_0, 即在近似显著性水平0.10 下可以认为原料产地的改变对该厂产品的质量没有发生显著的影响.

§6.3 拟合优度检验

前面的假设检验基本上是在假定总体是正态的条件下做的, 但是这个假设本身不一定成立, 需要收集样本$(X_1, \cdots, X_n)$ 来检验它. 一般地, 检验

$$H_0 : X\text{服从某种分布}$$

可以采用Karl Pearson 提出的χ^2 拟合优度检验.

§6.3.1 离散总体情形

(1) 理论分布不含未知参数的情形

设某总体X 服从一个离散分布, 且根据经验得知总体落在类别$a_1, \cdots, a_k$ 的理论频率分别为$p_1, \cdots, p_k$, 现从该总体抽得一个样本量为n 的样本, 其落在类别$a_1, \cdots, a_k$ 的观测数分别为$n_1, \cdots, n_k$. 感兴趣的问题是检验理论频率是否正确, 即下面假设是否正确:

$$H_0 : P(X \in a_1) = p_1, \cdots, P(X \in a_k) = p_k.$$

这类问题只提零假设而不提对立假设, 相应的检验方法称为拟合优度检验. 显然, 在零假设下, 各类别的理论频数分别为$np_1, \cdots, np_k$, 将理论频数和观测频数列于下表:

类别	a_1	a_2	$\cdots$	a_k
理论频数	np_1	np_2	$\cdots$	np_k
观测频数	n_1	n_2	$\cdots$	n_k

由大数定律知, 在零假设成立时, n_i/n 依概率收敛于p_i, 故理论频数np_i 与观测频数n_i 接近. 而检验统计量取为

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}.$$

简单地, 就是

$$\chi^2 = \sum \frac{(O-E)^2}{E},$$

其中O 为观测频数, E 为期望频数.

这个统计量中每项的分母的选取有点讲究, 我们可以这样粗略地解释: 假设n_i 服从Poisson 分布, 则n_i 的均值和方差均为np_i, 从而$(n_i - np_i)/\sqrt{np_i}$ 的极限分布为标准正态分布, 因此χ^2 近似为k 个服从自由度为1 的χ^2 分布的随机变量之和, 由于$\sum_{i=1}^{k}(n_i - np_i) = 0$, 故这$k$ 个随机变量满足一个约束, 从而χ^2 的自由度为$k-1$. 事实上, 可以严格地证明, 在一定的条件下, χ^2 的极限分布就是自由度为$k-1$ 的χ^2 分布, 但其证明超出本课程的要求范围.

下面给出一个例子来说明拟合优度检验的应用.

例 6.3.1. *有人制造一个含6 个面的骰子, 并声称是均匀的. 现设计一个实验来检验此命题: 连续投掷600 次, 发现出现六面的频数分别为97, 104, 82, 110, 93, 114. 问能否在显著性水平0.2 下认为骰子是均匀的?*

解: 该问题设计的总体是一个有6 个类别的离散总体, 记出现六个面的概率分别为$p_1, \cdots, p_6$, 则零假设可以表示为

$$H_0 : p_i = 1/6, i = 1, \cdots, 6.$$

在零假设下, 理论频数都是100, 故检验统计量χ^2 的取值为

$$\frac{(97-100)^2}{100} + \frac{(104-100)^2}{100} + \frac{(82-100)^2}{100} + \frac{(110-100)^2}{100} + \frac{(93-100)^2}{100} + \frac{(114-100)^2}{100} = 6.94,$$

跟自由度为$6-1=5$ 的χ^2 分布的上0.05 分位数$\chi_5^2(0.2) \approx 7.29$ 比较, 不能拒绝零假设, 即可在显著性水平0.2 下认为骰子是均匀的.

(2) 理论分布含若干未知参数的情形

当理论总体总含有未知的参数时, 理论频数np_i 一般也与这些参数有关, 此时应该用适当的估计如极大似然估计代替这些参数以得到p_i 的估计$\hat{p}_i$, 得到的统计量记为

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}.$$

拟合优度检验的提出者Karl Pearson 最初认为在零假设下, 检验统计量的χ^2 的极限分布仍等于自由度为$k-1$ 的χ^2 分布, R. A. Fisher 发现自由度应该等于$k-1$ 减去估计的独立参数的个数r, 即$k-1-r$.

例 6.3.2. *从某人群中随机抽取100 个人的血液, 并测定他们在某基因位点处的基因型. 假设该位点只有两个等位基因A 和a, 这100 个基因型中AA, Aa 和aa 的个数分别为30, 40, 30, 则能否在0.05 的水平下认为该群体在此位点处达到Hardy-Weinberg 平衡态?*

解: 取零假设为

$$H_0 : \text{Hardy-Weinberg 平衡态成立.}$$

设人群中等位基因A 的频率为p, 则该人群在此位点处达到Hardy-Weinberg 平衡态指的是在人群中3 个基因型的频率分别为$P(AA) = p^2$, $P(Aa) = 2p(1-p)$ 和$P(aa) = (1-p)^2$, 即零假设可等价地写成

$$H_0 : P(AA) = p^2, P(Aa) = 2p(1-p), P(aa) = (1-p)^2.$$

在H_0 下, 3 个基因型的理论频数为$100 \times \hat{p}^2$, $100 \times 2 \times \hat{p}^2(1-\hat{p})$ 和$100 \times (1-\hat{p})^2$, 其中$\hat{p}$ 等于估计的等位基因频率0.5, 代入χ^2 统计量表达式, 得统计量的值等于4. 该统计量的值大于自由度为$3-1-1=1$ (恰好一个自由参数被估计) 的χ^2 分布上0.05 分位数3.84, 故可在0.05 的水平下认为未达到Hardy-Weinberg 平衡态.

§6.3.2 列联表的独立性和齐一性检验

(1) 独立性检验

下面考虑很常用的列联表. 列联表是一种按两个属性作双向分类的表. 例如肝癌病人可以按所在医院(属性A) 和是否最终死亡(属性B) 分类. 目的是看不同医院的疗效是否不同. 又如婴儿可按喂养方式(属性A, 分两个水平: 母乳喂养与人工喂养) 和小儿牙齿发育状况(属性B, 分两个水平: 正常与异常) 来分类. 这两个例子中两个属性都只有两个

水平, 相应的列联表称为"四格表", 一般地, 如果第一个属性有a 个水平, 第二个属性有b 个水平, 称为$a\times b$ 表(见教材p268) . 实际应用中, 常见的一个问题是考察两个属性是否独立. 即零假设是

$$H_0 : \text{属性A 与属性B 独立}.$$

这是列联表的独立性检验问题.

假设样本量为n, 第(i,j) 格的频数为n_{ij}. 记$p_{ij}=P(\text{属性A, B 分别处于水平}i,j)$, $u_i=P(\text{属性A 有水平}i)$, $v_i=P(\text{属性B 有水平}j)$. 则零假设就是$p_{ij}=u_iv_j$. 将u_i 和v_j 看成参数, 则总的独立参数有$a-1+b-1=a+b-2$ 个. 它们的极大似然估计为

$$\hat{u}_i=\frac{n_{i\cdot}}{n}, \hat{v}_j=\frac{n_{\cdot j}}{n}.$$

正好是它们的频率(证明参看教材) . 其中$n_{i\cdot}=\sum_{j=1}^{b}n_{ij}$, $n_{\cdot j}=\sum_{i=1}^{a}n_{ij}$. 在$H_0$ 下, 第(i,j) 格的理论频数为$n\hat{p}_{ij}=n_{i\cdot}n_{\cdot j}/n$, 因此在$H_0$ 下, $\sum_{i=1}^{a}\sum_{j=1}^{b}(n_{ij}-n\hat{p}_{ij})$ 应该较小. 故取检验统计量为

$$\chi^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n)^2}{(n_{i\cdot}n_{\cdot j}/n)}.$$

在零假设下χ^2 的极限分布是有自由度为$k-1-r=ab-1-(a+b-2)=(a-1)(b-1)$ 的χ^2 分布. 对于四格表, 自由度为1.

(2) 齐一性检验

跟列联表有关的另一类重要的检验是齐一性检验, 即检验某一个属性A 的各个水平对应的另一个属性B 的分布全部相同, 这种检验跟独立性检验有着本质的区别. 独立性问题中两属性都是随机的; 而齐一性问题中属性A 是非随机的, 这样涉及到的分布实际上是条件分布. 虽然如此, 所采用的检验方法跟独立性检验完全一样.

例 6.3.3. *下面表是甲乙两医院肝癌病人生存情况. 需要根据这些数据判断两医院的治疗效果是否一样.*

甲、乙两院肝癌的近期疗效

	生存	死亡	合计
甲院	150(n_{11})	88(n_{12})	238($n_{1\cdot}$)
乙院	36(n_{21})	18(n_{22})	54($n_{2\cdot}$)
合计	186($n_{\cdot 1}$)	106($n_{\cdot 2}$)	292(n)

解: 这是一个齐一性检验问题. 检验统计量χ^2 的观测值为0.1195, 远远小于自由度为1 的χ^2 分布的上0.05 分位数, 故可以接受零假设, 即在水平0.05 下可以认为两个医院的疗效无差别的.

当有某个格子的频数较小时, 如果允许的话可以合并格子是每个格子的频数足够大, 实际问题中不允许合并格子(合并后失去了实际意义), 此时可以用**Fisher 的精确检验法**.

§6.3.3 连续总体情形

设$(X_1,\cdots,X_n)$ 是取自总体X 的一个样本, 记X 的分布函数为$F(x)$, 需要检验的那种分布中含有r 个总体参数$\theta_1,\cdots,\theta_r$. 我们要在显著性水平α 下检验

$$H_0: F(x) = F_0(x;\theta_1,\cdots,\theta_r),$$

其中$F_0(x;\theta_1,\cdots,\theta_r)$ 表示需要检验的那种分布的分布函数. 例如, 当我们要检验

$$H_0: X \sim N(\mu,\sigma^2)$$

时, $r=2$, $\theta_1=\mu, \theta_2=\sigma^2$.

$$F_0(x;\mu,\sigma^2) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\} dt.$$

上述假设可以通过适当的离散化总体分布, 采用拟合优度法来做检验. 首先把实数轴分成k 个子区间$(a_{j-1},a_j]$, $j=1,\cdots,k$, 其中a_0 可以取$-\infty$, a_k 可以取∞. 这样构造了一个离散总体, 其取值就是这k 个区间. 记

$$p_j \quad = \quad P_{H_0}(a_{j-1} < X \le a_j) = F_0(a_j;\theta_1,\cdots,\theta_r) - F_0(a_{j-1};\theta_1,\cdots,\theta_r), j=1,\cdots,k.$$

如果H_0 成立, 则概率p_j 应该与数据落在区间$(a_{j-1},a_j]$ 的频率$f_j = n_j/n$ 接近, 其中n_j 表示相应的频数. 当p_i 的取值不含未知参数时, 取检验统计量

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j},$$

否则取

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j},$$

其中$\hat{p}_i$ 是将p_i 中的未知参数换成适当的估计后得到的p_i 的估计. 拒绝域取为

$$\left\{\chi^2 > \chi^2_{k-r-1}(\alpha)\right\}.$$

如果p_i 中不含未知参数, 则$r = 0$.

使用χ^2 进行拟合优度检验时一般要求$n \geq 50$, $n\hat{p}_j \geq 5, j = 1, \cdots, k$, 如果不满足这个条件, 最好把某些组作适当合并.

例 6.3.4. *从某连续总体中抽取一个样本量为100 的样本, 发现样本均值和样本标准差分别为-0.225 和1.282, 落在不同区间的频数如下表所示:*

区间	$(-\infty, -1)$	$[-1, -0.5)$	$[-0.5, 0)$	$[0, 0.5)$	$[0.5, 1)$	$[1, \infty)$
观测频数	*25*	*10*	*18*	*24*	*10*	*13*
理论频数	*27*	*14*	*15*	*14*	*13*	*17*

可否在显著性水平0.05 下认为该总体服从正态分布?

解: 设理论正态分布的均值和方差分别为μ 和σ^2, 记第i 个区间为$(a_{i-1}, a_i, i = 1, \cdots, 6$, 则样本落在第$i$ 个格子的理论概数为$100P(a_{i-1} < X \leq a_i)$, 其中$X \sim N(\mu, \sigma^2)$. 将$\mu = -0.225$ 和$\sigma = 1.282$ 代入得到估计的理论频数, 列于上表中.

$$H_0: \text{总体服从正态分布}$$

由此算得检验统计量χ^2 的值约为9.34, 与自由度为5 的χ^2 分布的上0.1 分位数$\chi^2_5(0.1) \approx 9.24$ 比较可以拒绝零假设, 即可以在显著性水平0.1 下认为该总体不服从正态分布.