第一章：事件及其概率

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第一章事件与概率

1.1 概率论发展简史

♠ French Society in the 1650′ s

- 赌博是一种流行时尚
- 法律没有禁止
- 当游戏越来越复杂，利益越来越大时，需要数学方法来计算获胜可能性

♠ Enter the Mathematicians

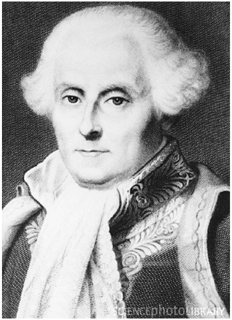
- 贵族**De Mere**在与一名宫廷卫士一次赌博时关于如何分赌本的问题发生了争执，于是请教他的好友数学家 **Blaise Pascal**.
- **Pascal** 与他的另一名好友数学家**Pierre Fermat**通信讨论了该问题，形成概率论中一个重要的基本概念 —数学期望

♠ Enter the Mathematicians

- **Christiaan Huygens** 在 1657 年写了世界上第一本关于概率论的著作 **De ratiociniis in ludo aleae (“ On Reasoning in Games of Chance”)**, 中文译名 “论赌博中的计算”

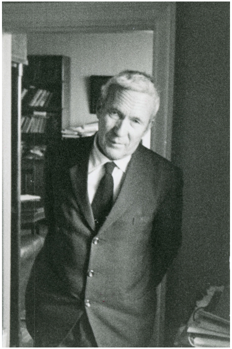
♠ Enter the Mathematicians

- **Pierre-Simon Laplace** 在他 1812 年的著作 **Theorie Analytique des Probabilities**中介绍了概率的数学理论及其科学应用.
- **Laplace**只考虑了古典概型, 对一般的概率及其应用没有介绍.
- 到 1850 年, 许多数学家发现古典概型对一般场合不合理, 开始尝试重新定义概率



♠ Enter the Mathematicians

- **Andrey Kolmogorov** 第一个在他 1933 年的著作 **Grundbegriffe der Wahrscheinlichkeitsrechnun**中严密的定义了概率.
- 类似于**Euclid**基于公理体系建立几何, 他从基本公理建立了概率理论, 从而使概率论称为一门严谨的数学分支
- 概率理论的现代研究和测度论非常紧密的结合在一起



概率论与数理统计的应用 概率统计理论与方法的应用几乎遍及所有科学技术领域、工农业生产和国民经济的各个部门中. 例如

- 气象、水文、地震预报、人口控制及预测都与《**概率论**》紧密相关
- 产品的抽样验收，新研制的药品能否在临床中应用，均要用到《**假设检验**》；
- 寻求最佳生产方案要进行《**实验设计**》和 《**数据处理**》；
- 电子系统的设计, 火箭卫星的研制及其发射都离不开《**可靠性估计**》
- 处理通信问题, 需要研究《**信息论**》；
- 研究经济数据等依时间观测数据时,《**时间序列分析**》 方法非常有用
- 研究化学反应的时变率，要以《**马尔可夫过程**》 来描述;

- 生物学中研究群体的增长问题时，提出了生灭型 **《随机模型》**，传染病流行问题要用到多变量非线性 **《生灭过程》**
- 许多服务系统，如电话通信、船舶装卸、机器维修、病人候诊、存货控制、水库调度、购物排队、红绿灯转换等，都可用一类概率模型来描述，其涉及到的知识就是 **《排队论》**
- 研究收入如何受个体教育程度, 行业, 性别等因素影响, 要用到 **《回归分析》**
- 研究多变量之间关系, 分类, 聚类等, 要用到 **《多元分析》**
- 研究寿命数据要用到 **《生存分析》**
- ...

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: 自然界中的一种客观现象, 当人们观测它时, 不能预先确定会出现哪种结果, 而仅仅知道是多种可能结果之一.

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, 在试验或观测之前不能预知是哪个结果发生
- 一般还要求试验在相同条件下能够重复

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.

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.

Definition

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.

随机事件：简称事件 (Event)，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.

Definition

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.

样本空间(Sample Space)： 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. 样本空间中的元素，称为样本点，通常用 ω 等表示.

Definition

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

(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.

↑Example

↓Example

- 样本空间的元素应该是相互不同的, 根据试验的不同目的, 样本空间应该予以不同的选择.
- 但是总的原则是样本空间应该尽可能详细, 即尽可能包含所有可能的结果.

看下面的例子

(1). 将一枚硬币抛三次，考察正反面出现的情况;
(2). 将一枚硬币抛三次，考察正面出现的次数。

↑Example

↓Example

这两个试验的目的不同，因此样本空间的选取也不同。

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.

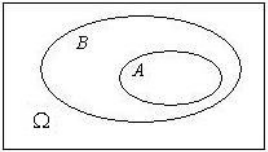
Definition

因此, 我们不严格的说**样本空间的子集称为随机事件**.

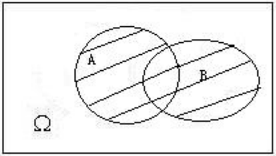
1.2.2 事件的关系及其运算

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.

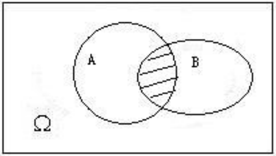
- 子事件 $A \subset B$: 事件 A 发生蕴含事件 B 一定发生, 则事件 A 称为事件 B 的子事件, 记为 $A \subset B$. 若 $A \subset B$, 且 $B \subset A$, 则称事件 A 与事件 B 相等, 记为 $A = B$.



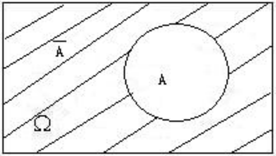
- 事件的和 $(A \cup B)$: 事件 A 和事件 B 中至少有一个发生的这一事件称为事件 A 和事件 B 的和, 记为 $A \cup B$.



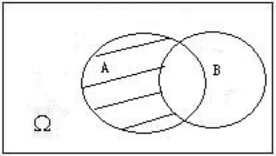
- 事件的积 $(A \cap B)$: 事件 A 和事件 B 同时发生这一事件称为事件 A 和事件 B 的积, 记为 $A \cap B$. 如果 $A \cap B = \phi$, 则称 A 和 B 不相容或者互斥, 即事件 A 和 B 不能同时发生.



- 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) .



- 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.



事件的运算

- 集合的运算法则适用于事件的运算
- De Morgan 对偶法则:

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

$\uparrow$Example

设 A, B, C 是三个事件，试表示下列事件

1. 事件 A, B 发生而 C 不发生；$(AB\bar{C})$
2. 事件 A, B, C 不同时发生; $(\bar{A} + \bar{B} + \bar{C})$
3. 事件 A, B, C 中至多有一个发生；$(A^cB^c + A^cC^c + B^cC^c)$
4. 事件 A, B, C 中至少发生两个；$(AB + AC + BC)$
5. 事件 A, B, C 中恰好发生两个；$(AB\bar{C} + A\bar{B}C + \bar{A}BC)$

$\downarrow$Example

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，其值习惯上用 0 和 1 之间的数表示，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，则事件 A 的概率为

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

记号：为方便起见，以 $|B|$ 记事件 B 中基本事件的个数．

(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.

↑Example

抛硬币的试验

试验者	掷硬币的次数	正面出现的次数	频率
蒲丰	4040	2048	.5069
皮尔逊	12000	6019	.5016
皮尔逊	24000	12012	.5005

↓Example

从这个例子可以看出随着试验次数的增加，频率越来越接近 1/2.

(3) 主观概率

- 个人对某个结果发生可能性的一个判断, 如某人认为有 80% 的可能性房价暴跌. 另一人则认为仅有 20% 的可能性.
- 有相当的生活基础
- 在金融和管理等方面有大量应用
- 基于此的概率学派称为贝叶斯 (Bayes) 学派
- 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,

称 $P(\cdot)$ 为一概率, 如果

(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 $A_1, A_2, \ldots$ 为两两不相容的事件序列, 则

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

Definition

注: 事实上我们把可以定义 (计算) 概率的事件集合记为 $\mathcal{F}$, 其是样本空间的 $\sigma-$ 代数.

由概率的公理化定义, 我们可以得到

1. $P(\phi)=0$

2. (有限可加性) 若 $A_k\in\mathcal{F}, k=1,\cdots,n$ 且两两互斥，则

$$P(\sum_{k=1}^{n}A_k)=\sum_{k=1}^{n}P(A_k)$$

3. (可减性) 若 $A,B\in\mathcal{F}$ 且 $A\subset B$，则 $P(B-A)=P(B)-P(A)$.

4. (单调性) 若 $A,B\in\mathcal{F}$ 且 $A\subset B$，则 $P(A)\leq P(B)$.

5. $P(\bar{A})=1-P(A)$

6. (加法定理) 对任意的事件 $A_1,\cdots,A_n\in\mathcal{F}$，有

$$\begin{aligned}P(\sum_{k=1}^{n}A_k) &= \sum_{k=1}^{n}P(A_k)-\sum_{1\leq i<j\leq n}^{n}P(A_iA_j)+\sum_{1\leq i<j<k\leq n}^{n}P(A_iA_jA_k)\\ &\quad -\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)\end{aligned}$$

7. (次可加性) 对任意的事件 $A_1, \cdots, A_n, \cdots \in \mathcal{F}$, 有 $P(\sum\limits_{n=1}^{\infty} A_n) \leq \sum\limits_{n=1}^{\infty} P(A_n)$.

8. (下连续性) 若 $A_n \in \mathcal{F}$ 且 $A_n \subset A_{n+1}, n = 1, 2, \cdots$, 则

$$P(\sum_{n=1}^{\infty} A_n) = \lim_{n} P(A_n)$$

9. (上连续性) 若 $A_n \in \mathcal{F}$ 且 $A_n \supset A_{n+1}, n = 1, 2, \cdots$, 则

$$P(\prod_{n=1}^{\infty} A_n) = \lim_{n} P(A_n)$$

(1992 年考研) 已知 $P(A) = P(B) = P(C) = 1/4$, $P(AB) = 0$, $P(AC) = P(BC) = 1/6$, 求 $P(\bar{A}\bar{B}\bar{C})$.

↑Example

↓Example

解

$$
\begin{aligned}
&P(\bar{A}\bar{B}\bar{C}) = 1 - P(A \cup B \cup C) \\
&= 1 - [P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)] \\
&= 1 - 3/4 + 2/6 = 7/12.
\end{aligned}
$$

但是又由 $P(A \cup B) = P(A) + P(B) - P(AB) = 1/2$ 可得

$$P(A \cup B \cup C) = 5/12$$

于是的矛盾

$$P(A \cup B) = 1/2 > 5/12 = P(A \cup B \cup C)$$

↑Example

求证对任意 n 个事件 $A_1, \cdots, A_n$ 有

$$P(\prod_{k=1}^{n} A_k) \geq \sum_{k=1}^{n} P(A_k) - n + 1$$

↓Example

第一章：事件及其概率

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 `http://staff.ustc.edu.cn/~cyu/`

论坛 `http://fisher.stat.ustc.edu.cn`

第一章事件与概率

1.3 古典概型

称一个随机试验为**古典概型**, 如果

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

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

记号: 为方便起见, 以 $|A|$ 记事件 A 中基本事件的个数.

计算古典概率, 主要用到排列组合的知识.

计数原理

乘法原理 假定进行过程 I 有 n_1 中方式，而对于过程 I 的每一个方式，进行过程 II 都有 n_2 种方式。那么，依次进行过程 I 与 II 共有 $n_1 n_2$ 种方式。

加法原理 假定进行过程 I 有 n_1 中方式，进行过程 II 有 n_2 种方式。那么，进行过程 I 或 II 共有 $n_1 + n_2$ 种方式。

排列组合

- 从 n 个不同的元素中, 有放回地取出 r 个元素组成的可重复排列的种数为 n^r 种。从 n 个不同的元素中，不放回地取出 r 个元素组成的不重复排列的种数为 $n(n-1)\cdots(n-r+1) = P_n^r$.

- 从 n 个不同的元素中, 不放回地取 r 个组成的组合，种数为

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

- 从 n 个不同的元素中, 有放回地取 r 个组成的组合 (不考虑顺序), 种数为

$$\binom{n+r-1}{r}$$

在运用排列组合公式时, 要清楚次序问题.

甲乙丙丁四人进行乒乓球双打练习, 两人一对地结为对打的双方, 有多少种不同的结对方式?

↑Example

↓Example

可能有人会认为这个问题是简单的组合问题: 从四人中选出两人结为一对, 剩下的两人结为一对即可. 于是他们算得: 有 $C_4^2 = 6$ 种方式.

但事实是否如此呢? 注意此时并不要求两队之间的顺序, 所以一共只有 3 种结对方式: $C_4^2/2 = 3$

“组合”是一种“有编号的分组模式”, 或者说, 按照组合模式计算出的分组方式数目中, 已经天然地把组的不同编号方式数目计算在内了.

欲将 6 个人分为 3 组, 每组 2 人, 分别从事 3 项不同工作, 求分配方式数.

↑Example

↓Example

解: 先取出两人从事第 1 项工作, 有 C_6^2 种方式; 再取出两人从事第 2 项工作, 有 C_4^2 种方式; 剩下的两人从事第 3 项工作. 所以一共有:

$$C_6^2 \cdot C_4^2 = \frac{6!}{4! \cdot 2!} \frac{4!}{2! \cdot 2!} = \frac{6!}{2! \cdot 2! \cdot 2!} = 90$$

种分配方式.

要把 7 人分为 3 个小组, 执行同一种任务, 其中一个组 3 人, 另两个组各 2 人, 求分组方式数.

↑Example ↓Example

解: 显然这也是一个 "无编号分组" 问题. 但是却与上面的情况有所不同. 因为其中有一个 3 人组, 无论是否编号, 它都与其余两个组有所区别 (编号无非是为了对分出的组加以区分), 所以在按 "有编号分组模式" 算出分组方式数之后, 只应再除以 2! (即除去两个不加区分的组的排列顺序数), 故得: 共有

$$\frac{7!}{3!\cdot 2!\cdot 2!}\cdot\frac{1}{2!}=\frac{7!}{3!\cdot(2!)^3}$$

种分组方式.

为了适应这种分为多个 "不同的" 组的问题需求, 人们总结出如下的 "多组组合模式":

多组组合模式： 有 n 个不同元素, 要把它们分为 k 个不同的组, 使得各组依次有 $n_1, n_2, \cdots, n_k$ 个元素, 其中 $n_1+n_2+\cdots+n_k=n$, 则一共有

$$\frac{n!}{n_1!\cdot n_2!\cdot \ldots \cdot n_k!}$$

种不同分法.

不尽相异元素的排列模式 有 n 个元素, 属于 k 个不同的类, 同类元素之间不可辨认, 各类元素分别有 n_1, n_2, $\cdots$, n_k 个, 其中 $n_1+n_2+\cdots+n_k=n$, 要把它们排成一列, 则一共有

$$\frac{n!}{n_1!\cdot n_2!\cdots n_k!}$$

种不同排法.

一批产品有 N 个，其中废品有 M 个。现从中随机取出 n 个，在以下两种情形下，分别求“其中恰好有 m 个废品”这一事件的概率。(1) 有放回地选取； (2) 不放回地选取

↑Example

↓Example

解: 记 $A=\{$其中恰好有 m 个废品$\}$，则

(1) 有放回情形 $|\Omega|=N^n$， $|A|=\binom{n}{m}M^m(N-M)^{n-m}$, 所以

$$P(A)=\frac{C_n^m M^m(N-M)^{n-m}}{N^n}=\begin{pmatrix} n \\ m \end{pmatrix}\left(\frac{M}{N}\right)^m\left(\frac{N-M}{N}\right)^{n-m}$$

(2) 不放回情形 $|\Omega|=C_N^n$， $|A|=C_M^m C_{N-M}^{n-m}$, 所以

$$P(A)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$$

n 个男生, m 个女生排成一排 $(m \leq n+1)$. 求事件 $A = $ {任意两个女孩不相邻} 的概率。又若排成一圈，又如何?

↑Example ↓Example

解: (1) 排成一排

$$|\Omega| = (n+m)!, \quad |A| = n!C_{n+1}^{m}m!$$
$$P(A) = \frac{|A|}{|\Omega|} = \frac{n!C_{n+1}^{m}m!}{(n+m)!}$$

(2) 排成一圈

$$|\Omega| = (n+m-1)!, \quad |A| = (n-1)!C_{n}^{m}m!$$
$$P(A) = \frac{|A|}{|\Omega|} = \frac{(n-1)!C_{n}^{m}m!}{(n+m-1)!}$$

↑Example

r 个不同的球任意放入编号为 1 至 n 的 n 个盒子，每球入各盒均等可能，求下列事件的概率

(1) A={指定的 r 个盒子各含一个球}

(2) B={每盒至多有一球}

(3) C={某指定盒中恰有 m 个球}

↓Example

解: $|\Omega| = n^r$

(1) $|A| = r!$

(2) $|B| = C_n^r r!$

(3) $|C| = C_r^m (n-1)^{r-m}$

又若球是相同的，则在这里,r 个球是相同的,n 个盒子是互不相同的. 因此我们只需要关心各个盒子中的球数, 而无需考虑哪个球落在哪个盒子中. 我们可把问题设想为:

r 个相同的小球已经一字排开, 只须在它们之间加上 $n-1$ 块隔板, 把它们隔为 n 段, 然后让各段对号放入相应的盒子即可. 由于盒子可空, 相当于要将 $r+n-1$ 个不尽相异的元素进行排列, 其中 1 类

元素 (小球) 有 r 个, 另 1 类元素 (隔板) 有 $n-1$ 个, 所以由不尽相异元素的排列模式知, 一共有

$$C_{r+n-1}^{r} = C_{r+n-1}^{n-1}$$

种不同分法. 因此

$$|\Omega| = \binom{n+r-1}{n-1}$$

(1) $|A| = 1$

(2) $|B| = C_n^r$

(3) $|C| = \binom{r-m+n-1-1}{r-m}$

注：

- 球相异和球相同两种情形下的样本空间是不同的，即机会均等原则是不同的。(各是什么呢？)

- 这个例子是古典概型中一个很典型的问题，不少实际问题可以归结为它

- 若把球解释为粒子，把盒子解释为相空间中的小区域，则这个问题便相应于统计物理学里的 Maxwell—Boltzmann 统计

- **生日问题**

求 r 个人中没有两个人生日相同的概率．若把 r 个人看作上面问题中的 r 个球，而把一年的 365 看作为盒子，则 $n = 365$，这时事件 B 的概率即为所求概率。例如当 $r = 40$ 时，$P(B) = 0.109$，这个概率已经相当小；而当 $r = 50$ 时, $P(B) = 0.03$。进一步当 $r = 55$ 时，$P(B)$ 之值只有 0.01，这实在是出乎意料地小。

设有方程 $x+y+z=15$, 试分别求出它的正整数解和非负整数解 (x,y,z) 的组数.

解: 本题可以设想为将 15 个无区别的小球分入 3 个不同的盒子, 再分别将第 $1,2,3$ 个盒中的球数对应为 x,y,z 的值即可. 所以, 非负整数解的组数 (相当于允许出现空盒的情况) 为:

$$C_{15+3-1}^{15}=C_{17}^{2}=\frac{17\times 16}{2}=136;$$

而正整数解的组数 (相当于不允许出现空盒的情况) 为:

$$C_{15-1}^{3-1}=C_{14}^{2}=\frac{14\cdot 13}{2}=91. \quad \#$$

注: 此例的方法即是证明前面排列组合的方法。

设有 n 个人随机地坐到礼堂第一排 N 个座位上去，试求下列事件的概率：(1) 任何人都没有邻座；(2) 每人恰有一个邻座；(3) 关于中央座位对称的两个座位至少有一个空着。

↑Example

↓Example

解: 分别用 A, B, C 表示上述 (1)-(3) 各事件。则 $|\Omega| = P_N^n$.

(1) 视此 n 个人为“女生”，$N-n$ 个座位为“男生”，则 $|A| = C_{N-n+1}^n n!$

(2) $|B| = C_{N-n+1}^{n/2} n!$

(3)

$$|C| = \begin{cases} C_{N/2}^n 2^n n!, & N \ is \ even \\ nC_{(N-1)/2}^{n-1} 2^{n-1}(n-1)! + C_{(N-1)/2}^n 2^n n!, & N \ is \ odd \end{cases}$$

1.4 几何概型

在实际中，我们还会碰到样本点无限多的情形。几何概型是其中的一种.

设 Ω 是欧氏空间中确定的集合，满足条件 $0 < m(\Omega) < +\infty$。对 Ω 中的任何可测子集 A，称

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

为事件 A 的几何概率。这里等可能性体现在“落在区域 A 的概率与区域 A 的测度成正比并且与其形状位置无关。"

Definition

这里 $m(\Omega)$ 表示 Ω 的“大小”.

甲乙两人约定在 $[0, T]$ 时段内去某地会面，规定先到者等候一段时间 $t(t \leq T)$ 再离去。试求事件 A={甲乙将会面} 的概率。

↑Example

↓Example

解: 以 x, y 分别表示甲乙到达会面地点的时间。则

$$\Omega = \{(x, y)|0 \leq x, y \leq T\}$$

而 $A = \{(x, y)||x - y| \leq t\}$, 因此

$$P(A) = \frac{m(A)}{m(\Omega)} = 1 - (1 - \frac{t}{T})^2.$$

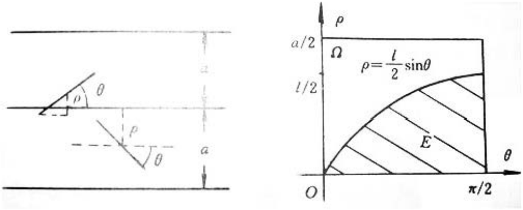
(Buffon's needle) 桌面上画满间隔均为 a 的平行线，现向桌面任意投放一长为 $l(l < a)$ 的针，求事件 E ={针与某直线相交} 的概率。

↑Example

↓Example

解：如下图所示，针的位置由针的中点到最近直线的距离 ρ 及针与

图 1.1: 针和平行线位置关系



直线所夹锐角 θ 决定。于是 $\Omega=\{(\rho,\theta):0\le\rho\le a/2,0\le\theta\le\pi/2\}$. 由针的任意性，样本点 (ρ,θ) 在 Ω 中均匀分布，是几何概型。而针与某直线相交，当且仅当 $\rho\le\frac{l}{2}sin\theta$. 即

$$E=\{(\rho,\theta)\in\Omega:\rho\le\frac{l}{2}sin\theta\}$$

$$m(\Omega)=\frac{\pi a}{4},\quad m(E)=\int_0^{\pi/2}\frac{l}{2}sin\theta d\theta$$

所以

$$P(E)=\frac{m(E)}{m(\Omega)}=\frac{2l}{\pi a}.$$

从而

$$\pi\approx\frac{2l}{\hat{p}a}$$

值得注意的是这里采用的方法：建立一个概率模型，它与某些我们感兴趣的量 —这里是常数 π—有关，然后设计随机试验，并通过这个试验的结果来确定这些量。这也就是 Monte-Carlo 思想.

在圆周上任取三点 A, B, C，求事件 E=$\{\triangle ABC$ 为锐角三角形$\}$的概率。(1/4)

↑Example ↓Example

在圆上任取两点 A, B 连成一条弦，再任取两点 C, D 连成一弦，求 AB 与 CD 相交的概率。(1/3)

↑Example ↓Example

第一章：事件及其概率

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第一章事件与概率

1.5　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%, 这个概率就是条件概率.

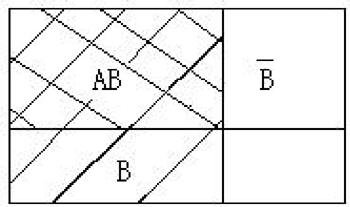
设事件 A 和 B 是随机试验 Ω 中的两个事件, $P(B) > 0$,
称

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

为事件 B 发生条件下事件 A 发生的条件概率.

Definition

注 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.



注 2. 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 3. 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随机试验看成无条件的，则在补充条件下进行的随机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．

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.

↑Example

↓Example

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, $P(B|A) = 2/9 \neq P(A) = 3/10$.

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:

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

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.

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.

↑Example ↓Example

解: 令 A_i={第 i 次打通电话}, $i=1,2,3$, 则

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

将 n 根短绳的 $2n$ 个端头任意两两连接, 试求恰好连成 n 个圈的概率.

↑Example

↓Example

解: 以 Ω 表示所有不同连结结果的集合, 设想把 $2n$ 个端头排成一行, 然后规定将第 $2k-1$ 个端头与第 $2k$ 个端头相连接,$k=1,2,\cdots,n$. 于是每一种排法对应一种连结结果, 从而 $|\Omega|=(2n)!$. 以 A 表示恰好连成 n 个圈的事件. 设想已将 n 根短绳作了编号, 以 A_k 表示第 k 号短绳被连成 1 个圈的事件, 于是有 $A=\bigcap\limits_{k=1}^{n} A_k$.

当 A_1 发生时,1 号短绳被连成 1 个圈, 这相当于有一个 $k\in\{1,2,\cdots,n\}$, 使得在 $2n$ 个端头的排列中,1 号短绳的两个端头排在第 $2k-1$ 和第 $2k$ 个位置上, 所以 $|A_1|=2n(2n-2)!$. 因此

$$P(A_1)=\frac{|A_1|}{|\Omega|}=\frac{1}{2n-1}\ .$$

我们来求 $P(A_2|A_1)$, 即要在已知 1 号短绳被连成 1 个圈的情况下, 求 2 号短绳也被连成 1 个圈的概率. 既然 1 号短绳已经自成 1 个圈, 我们就可以不考虑它, 只要对剩下的 $n-1$ 根短绳讨论其中的头一号短绳被连成 1 个圈的问题就行了. 即知

$$P(A_2|A_1) = \frac{1}{2(n-1)-1} = \frac{1}{2n-3} .$$

同理可得

$$P(A_k|A_1A_2\cdots A_{k-1}) = \frac{1}{2[n-(k-1)]-1} = \frac{1}{2n-2k+1}, \quad k = 3, 4, \cdots, n.$$

于是由概率乘法定理中的 (2.3.6) 式得到

$$P(A) = P(\bigcap_{k=1}^{n} A_k) = \prod_{k=1}^{n} \frac{1}{2n-2k+1} = \frac{1}{(2n-1)!!} .$$

在这个解法中, 充分体现了利用变化了的概率空间计算条件概率的好处.

1.5.1 全概率公式和 Bayes 公式

1. 全概率公式

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*）.

Definition

全概率公式:

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

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

$\overline{\uparrow}$Example

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 B_1 厂提供的, B_2 厂商和 B_3 分别提供 25% . 已知厂商 B_1 和 B_2 的次品率都是 2%, B_3 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.

$\underline{\downarrow}$Example

解: 记 A ={取出的产品为次品}, 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

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

将 n 根短绳的 $2n$ 个端头任意两两连接, 求恰好连成 n 个圈的概率.

↑Example ↓Example

解: 现在再来利用全概率公式给出一个解答. 以 A_n 表示 n 根短绳恰好连成 n 个圈的事件, 记 $p_n = P(A_n)$. 再以 B 表示第 1 根短绳连成 1 个圈的事件, 用 B 和 B^c 作为对 Ω 的一个分划. 于是由全概率公式得

$$p_n = P(A_n) = P(B)P(A_n|B) + P(B^c)P(A_n|B^c).$$

在前面例子中已经求得 $P(B) = \frac{1}{2n-1}$; 易见 $P(A_n|B^c) = 0$; 而 $P(A_n|B)$ 则是在已知第 1 根短绳连成 1 个圈的条件下, 其余 $n-1$ 根短绳连成 $n-1$ 个圈的概率, 此时第 1 根短绳已经与其余 $n-1$ 根

短绳无关, 所以 $P(A_n|B) = P(A_{n-1}) = p_{n-1}$, 代入上式即可得到

$$p_n = P(A_n) = \frac{1}{2n-1} p_{n-1} \ , \quad n = 2, 3, \cdots \ .$$

反复利用该式, 并注意 $p_1 = 1$, 即得

$$p_n = \frac{1}{(2n-1)!!} \ , \quad n = 1, 2, \cdots \ .$$

↑Example

(**Polya 罐子模型**) 罐中放有 a 个白球和 b 个黑球, 每次从罐中随机抽取一个球, 并连同 c 个同色球一起放回罐中, 如此反复进行. 试证明: 在第 n 次取球时取出白球的概率为 $\frac{a}{a+b}$.

↓Example

证: 以 A_k 表示在第 k 次取球时取出白球的事件, 于是 A_k^c 就是在第 k 次取球时取出黑球的事件. 我们来对 n 作归纳. 显然有 $P(A_1)=\frac{a}{a+b}$. 假设 $n=k-1,\ k\geq 2$ 时结论成立, 要证 $n=k$ 时结论也成立. 我们以 A_1 和 A_1^c 作为对 Ω 的一个分划. 注意此时可将 $P(A_k|A_1)$ 看成是从原来放有 $a+c$ 个白球和 b 个黑球的罐中按规则取球, 并且在第 $k-1$ 次取球时取出白球的概率, 因此由归纳假设知 $P(A_k|A_1)=\frac{a+c}{a+b+c}$, 同理亦有 $P(A_k|A_1^c)=\frac{a}{a+b+c}$, 于是由全概率公

式得

$$\begin{aligned} P(A_k) &= P(A_1)P(A_k|A_1)+P(A_1^c)P(A_k|A_1^c) \\ &= \frac{a}{a+b}\,\frac{a+b}{a+b+c}+\frac{b}{a+b+c}\,\frac{a}{a+b+c}=\frac{a}{a+b}\,. \end{aligned}$$

因此结论对一切 n 成立.

↑Example

一罐内有 a 个黑球和 b 个白球，从中任意取一球，如果是白球则将它放回去，如果是黑球，则从另一罐内取一白球替换它放回去。在重复 n 次这样的做法后，求第 n+1 次取出的是白球的概率。

↓Example

解: 记 A={第 n 次取出的是白球}，$p_n = P(A)$，B 为所求事件。则

$$
\begin{aligned}
p_{n+1} &= P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) \\
&= p_n * p_n + (p_n + \frac{1}{a+b})(1-p_n) \\
&= \left(1 - \frac{1}{a+b}\right)p_n + \frac{1}{a+b}
\end{aligned}
$$

结合初值 $p_1 = \frac{a}{a+b}$，得到

$$p_{n+1} = 1 - \left(1 - \frac{1}{a+b}\right)^n \frac{b}{a+b}.$$

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.

↑Example

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.

↓Example

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%.

1.5.2 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:

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. Definition

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.

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

Definition

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

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

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

若 $A_1, \cdots, A_n$ 中任意两个事件相互独立, 则称为**两两独立**.

注意: 独立和不相容是不同的两个概念.

↑Example

(两两独立而不相互独立的反例) 有四个同样的小球, 分别在其上写上“1”, “2”, “3”和“1,2,3”。引进三个事件: A_i={随机取一球, 球上有数字 i}, $i = 1, 2, 3$. 试讨论事件 A_1, A_2, A_3 是否相互独立.

↓Example

解: 易知 $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$, $P(A_1A_2) = P(A_2A_3) = P(A_3A_1) = \frac{1}{4}$, 但是却有 $P(A_1A_2A_3) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3)$, 所以事件 A_1, A_2, A_3 两两独立, 但不相互独立. 这个例子说明**两两独立不一定独立**。

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?

↑Example

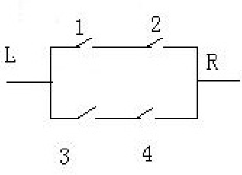
↓Example

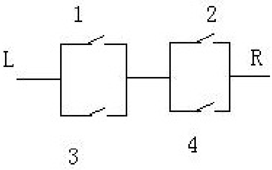
解: 设 D={密码被破译}, A, B 和 C 分别表示 A, B 和 C 三人能破译密码这三个事件, 由独立性,

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

↑Example

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.

↓Example

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.

假设某个人独立向同一目标射击 n 次, 每次命中目标的概率为 $p(p < 0.05)$, 求他在 n 次射击中至少有一次命中目标的概率.

↑Example

↓Example

解: 令 A_i ={第 i 次命中目标}, D ={至少有一次命中目标}, 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 较小时成立. 例如 $p = 0.04$, $n = 100$ 时, $P(D) \approx 1 - \exp\{-4\} = 0.98168$.

小概率原理

一个事件如果发生的概率很小的话，那么它在一次试验中是几乎不可能发生的，但在多次重复试验中几乎是必然发生的，数学上称之**小概率原理**。

↑Example

某停车场有 16 个并排的车位, 有天发现有 12 个车位停了车, 4 个相连的车位还空着. 由此能得出什么结论?

↓Example

- 购买彩票
- 乘坐飞机
- …

第二讲：随机变量及其分布

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第二章随机变量及其分布

1.1 随机变量的概念

随机变量是其值随机会而定的变量。

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.

↑Example

↓Example

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.

↑Example

↓Example

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.

↑Example ↓Example

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.

掷一枚硬币出现正面或反面.

↑Example ↓Example

产品被分为正品或废品.

↑Example ↓Example

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.

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.

Definition

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.

1.2 离散型随机变量

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.

Definition

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.

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

$$p_i = P(X = a_i), \ \ i = 1, 2, \ldots \tag{1.1}$$

称为 X 的概率质量函数 (probability mass function, *pmf*) 或分布律.

Definition

概率质量函数 $\{p_i,\ i=1,2,..\}$ 必须满足下列条件:

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

概率函数 (1.1) 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	...

(1.2)

有时也把 (1.2) 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 $x_1, x_2, \ldots$.. 令 A 为 $\{x_1, x_2, \ldots\}$ 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.

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.

Definition

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.

Definition

下面的 0-1 分布和二项分布都是以 Bernoulli 试验为基础的.

1.2.1 0-1 分布

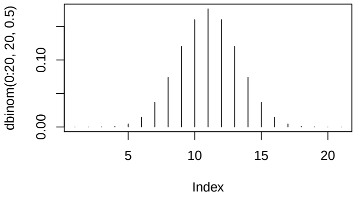
设随机变量 X 只取 0,1 两值，$P(X = 1) = p$，$P(X = 0) = 1 - p$，则称 X 服从 0-1 分布或 Bernoulli 分布. 0-1 分布是很多古典概率模型的基础.

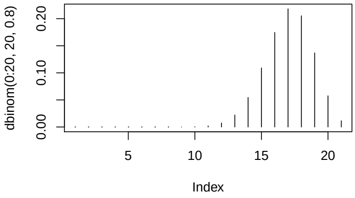
1.2.2 二项分布

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

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n. \tag{1.3}$$

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 (1.3) 确实是一个概率函数.

in R

```
dbinom, rbinom, pbinom, qbinom
```

↑Code ↓Code

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:

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

一个变量服从二项分布有两个条件:

- 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
- 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.

1.2.3 几何分布 (Geometric distribution)

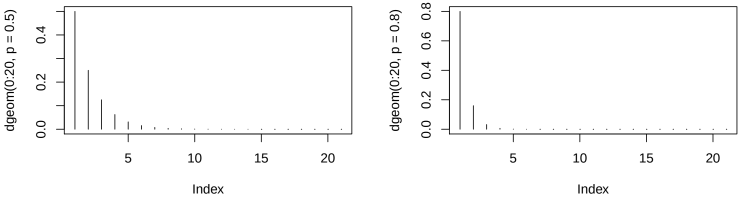
在 **n** 重贝努里实验中，当试验次数 $n \to \infty$ 时，称为可列重贝努里试验。

Definition

若以 X 表示在可列重贝努里试验中结果 A 出现时的试验次数，即若以“成功”表示结果 A 发生，则 X 表示首次成功时的试验次数，所以

$$P(X = k) = q^{k-1}p, \quad k = 1, 2, \cdots . \tag{1.4}$$

称此分布为几何分布. 记为 $X \sim G(p)$.



in R

```
dgeom, rgeom pgeom, qgeom
```

↑Code ↓Code

一个人要开门, 他共有 n 把钥匙。其中仅有一把可以打开门。现随机地有放回的从中选取一把开门，问这人在第 S 次试开成功的概

↑Example

率。

↓Example

定理 1. *以所有正整数为取值集合的随机变量 ξ 服从几何分布 $G(p)$, 当且仅当对任何正整数 m 和 n, 都有*

$$P(\xi > m + n \mid \xi > m) = P(\xi > n). \tag{1.5}$$

这个性质称为几何分布的无记忆性 (memoryless property).

证: 设随机变量 ξ 服从几何分布 $G(p)$, 写 $q = 1 - p$, 那么对任何非负整数 k, 都有

$$P(\xi > k) = \sum_{j=k+1}^{\infty} P(\xi = j) = p \sum_{j=k+1}^{\infty} q^{j-1} = q^k .$$

所以对任何正整数 m 和 n, 都有

$$P(\xi > m + n \mid \xi > m) = \frac{P(\xi > m + n,\ \xi > m)}{P(\xi > m)}$$

$$= \frac{P(\xi > m+n)}{P(\xi > m)} = \frac{q^{m+n}}{q^n} = q^n = P(\xi > n).$$

故知 (1.5) 式成立.

反之, 设对任何正整数 m 和 n, 都有 (1.5) 式成立. 对非负整数 k, 我们记 $p_k = P(\xi > k)$. 于是由 (1.5) 式知, 对任何正整数 k, 都有 $p_k > 0$, 并且对任何正整数 m 和 n, 都有 $p_{m+n} = p_m \cdot p_n$. 由此等式立知, 对任何正整数 m, 都有 $p_m = p_1^m$. 由于 $p_1 > 0$, 而若 $p_1 = 1$, 则必导致对一切正整数 m, 都有 $p_m = 1$, 此为不可能, 所以对某个小于 1 的正数 q, 有 $p_1 = q$. 由此不难得, 对任何正整数 m, 都有

$$P(\xi = m) = P(\xi > m-1) - P(\xi > m) = p_{m-1} - p_m = q^{m-1} - q^m = p\, q^{m-1},$$

其中 $p = 1 - q$, 所以 ξ 服从几何分布 $G(p)$.

1.2.4 Pascal 分布 (负二项分布)

在可列重贝努里试验中，若以 X_r 表示第 r 次成功发生时的试验次数，则 X_r 的分布律为

$$\begin{aligned} P(X_r=k) &= P(\{\text{前}k-1\text{次恰有}r\text{次成功且第}k\text{次成功}\}) \\ &= P(\{\text{前}k-1\text{次恰有}r\text{次成功}\})P(\{\text{第}k\text{次成功}\}) \\ &= C_{k-1}^{r-1}p^{r-1}q^{k-r}\cdot p \\ &= C_{k-1}^{r-1}p^{r}q^{k-r}, \quad k=r,r+1,\cdots. \end{aligned}$$

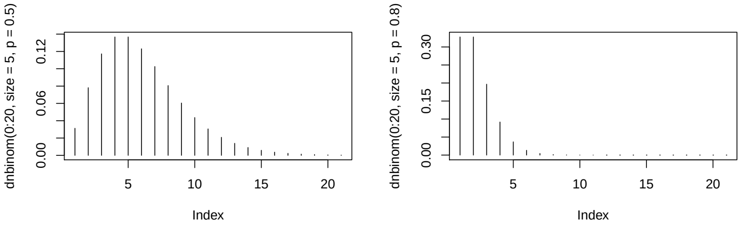
称此概率分布为 Pascal 分布。

in R

```
dnbinom, rnbinom, pnbinom, qnbinom
P(X_r = k)=dnbinom(k-r,size=r,prob=p)
```

↑Code

↓Code



如果记

$$p_k = C_{k-1}^{r-1} p^r q^{k-r}, \quad k = r, r+1, \cdots \tag{1.6}$$

那么显然有

$$\sum_{k=r}^{\infty} p_k = \sum_{k=r}^{\infty} C_{k-1}^{r-1} p^{\,r} q^{k-r} = p^{\,r} \sum_{k=0}^{\infty} C_{r+k-1}^{r-1} q^{\,k} = p^{\,r}(1-q)^{-r} = 1\ ,$$

所以 (1.6) 式的确是一个离散型随机变量的分布律. 我们将其称为参

数为 p 和 r 的 Pascal 分布. 又因为上式表明, 它可以用负二项展开式中的各项表示, 所以又称为负二项分布.

(Banach 火柴问题) 某人口袋里放有两盒火柴, 每盒装有火柴 n 根. 他每次随机取出一盒, 并从中拿出一根火柴使用. 试求他取出一盒, 发现已空, 而此时另一盒中尚余 r 根火柴的概率.

↑Example

↓Example

解: 以 A 表示甲盒已空, 而此时乙盒中尚余 r 根火柴的事件. 由对称性知, 所求的概率等于 $2P(A)$. 我们将每取出甲盒一次视为取得一次成功, 以 ξ 表示取得第 $n+1$ 次成功时的取盒次数, 则 ξ 服从参数为 0.5 和 $n+1$ 的 Pascal 分布 (因为每次取出甲盒的概率是 0.5). 易知, 事件 A 发生, 当且仅当 ξ 等于 $2n-r+1$. 所以所求的概率等于

$$2P(A) = 2P(\xi = 2n-r+1) = C_{2n-r}^{n} 2^{\,r-2n}.$$

↑Example

在可列重贝努里试验中，求事件 E ={n 次成功发生在 m 次失败之前} 的概率。

↓Example

解: 记 F_k={第 n 次成功发生在第 k 次试验}, 则

$$E = \bigcup_{k=n}^{n+m-1} F_k$$

且诸 F_k 两两互斥，故

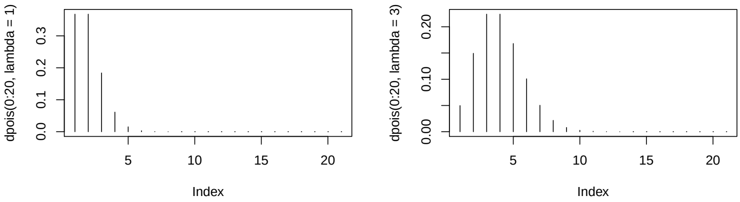
$$P(E) = \sum_{k=n}^{n+m-1} P(F_k) = \sum_{k=n}^{n+m-1} C_{k-1}^{n-1} p^n q^{n-k}.$$

1.2.5 Poisson 分布

设随机变量 X 的概率分布为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots\quad \lambda>0, \tag{1.7}$$

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

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

所以
$$\sum_{k=0}^{\infty} P(X=k)=1.$$

in R

```
dpois, rpois, ppois, qpois
```

↑Code ↓Code

↑Example

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?

↓Example

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

$$\binom{N}{x}\left(\frac{D}{V}\right)^{x}\left(1-\frac{D}{V}\right)^{N-x}. \tag{1.8}$$

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 (1.8) 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 (1.8) 式改写成如下形式:

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

当 N 变成无限时其极限为

$$e^{-Dd}(Dd)^{x}/x! \tag{1.9}$$

令 $Dd = \lambda$，则 (1.9) 和 (1.7) 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. *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，*

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}. \tag{1.10}$$

↑Example

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?

↓Example

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.

↑Example

假设一块放射性物质在单位时间内发射出的 α 粒子数 ξ 服从参数为 λ 的 Poisson 分布。而每个发射出来的 α 粒子被记录下来的概率是 p，就是说有 $q = 1 - p$ 的概率被记数器漏记。如果各粒子是否被记数器记录是相互独立的，试求记录下来的 α 粒子数 η 的分布。

↓Example

解: 以事件 $\{\xi = n\}, n = 0, 1, 2, \cdots$ 为划分，则由全概率公式有

$$
\begin{aligned}
P(\eta = k) &= \sum_{n=0}^{\infty} P(\eta = k | \xi = n) P(\xi = n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} \\
&= \sum_{n=k}^{\infty} \frac{(\lambda q)^{n-k}}{k!(n-k)!} e^{-\lambda} (\lambda p)^k = \frac{(\lambda p)^k}{k!} e^{-\lambda p}, \quad k = 0, 1, 2, \cdots .\#
\end{aligned}
$$

1.2.6 离散的均匀分布

设随机变量 X 取值 $a_1, a_2, ..., a_n$, 且有

$$P(X = a_k) = \frac{1}{n}, \ k = 1, ..., n. \tag{1.11}$$

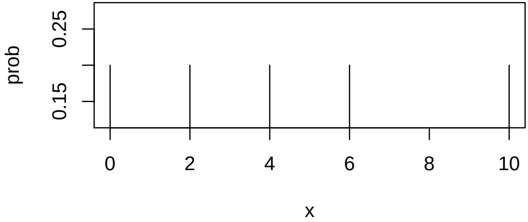
则称 X 服从离散的均匀分布.

可以看出, 离散的均匀分布正是古典概型的抽象.

in R

```
x<-c(0,2,4,6,10)
prob<-rep(0.2,5)
plot(x,prob,type="h")
```

↑Code ↓Code



第二讲：随机变量及其分布

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第二章随机变量及其分布

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:

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 $0 < X \leq 2$ 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 $-0.25 < X \leq 1.5$ 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:

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 $P(a \leq X \leq b) = \int_a^b f(x)dx.$

Definition

注 1. 对于任意的 $-\infty < x < +\infty$, 有 $P(X = x) = \int_x^x f(u)du = 0.$

注 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 3. 假设有总共一个单位的质量连续地分布在 $a\le x\le b$ 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.

设 X 为一连续型随机变量. 则

$$F(x) = \int_{-\infty}^{x} f(u)du, \quad -\infty < x < +\infty \tag{2.1}$$

Definition

称为 X 的 (累积) 分布函数.

注 4. $F(x)$ 表示的是随机变量的数值小于或等于 x 的概率, 即

$$F(x) = P(X \leq x) \quad -\infty < x < +\infty. \tag{2.2}$$

由式 (2.2) 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;

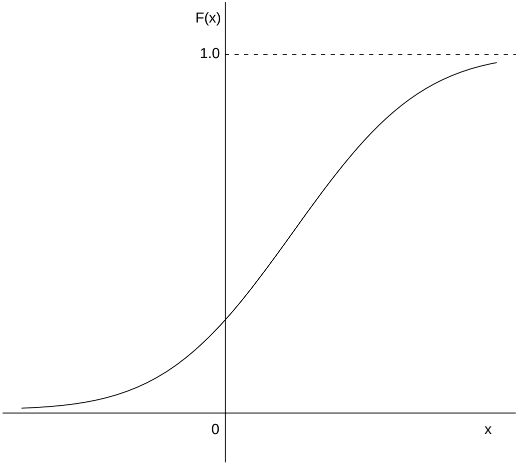
(2) $\lim_{x\to-\infty} F(x) = 0$;

(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.


F(x)
1.0
0
x

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.

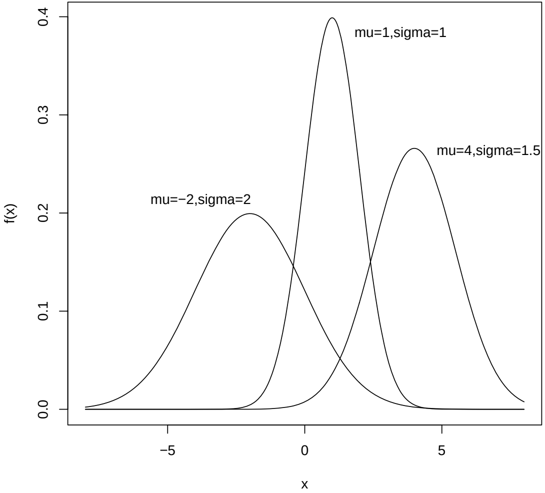
2.3.1 正态分布

如果一个随机变量 X 具有概率密度函数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < +\infty \tag{2.3}$$

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) 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.


mu=1,sigma=1
mu=4,sigma=1.5
mu=−2,sigma=2
f(x)
0.0
0.1
0.2
0.3
0.4
−5
0
5
x

in R

↑Code

```
x <- seq(-4, 4, length = 401)
plot(x, dnorm(x), type = 'l') # N(0, 1)
N(1, 1.5^2):
lines(x, dnorm(x, mean = 1, sd = 1.5), lty = 'dashed')
```

↓Code

从上图可以看出，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.

↑Example

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

$$P(\mu - k\sigma < x < \mu + k\sigma) = F(\mu + k\sigma) - F(\mu - k\sigma) = \Phi(k) - \Phi(-k) = 0.95. \quad (2.4)$$

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 $k = 1.96$.

↓Example

2.3.2 指数分布

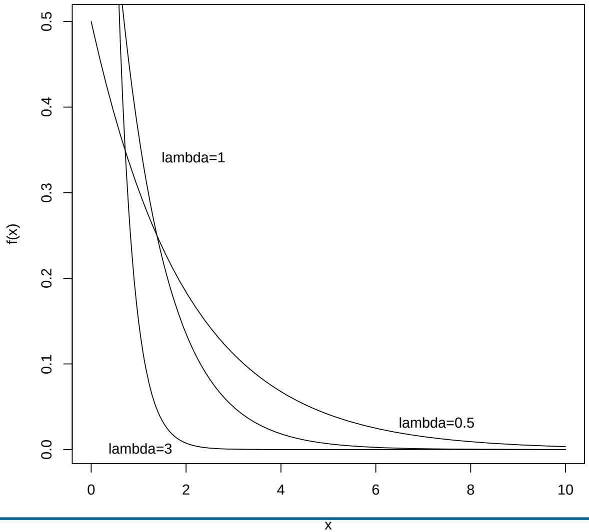
若随机变量 X 具有概率密度函数

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x \leq 0, \end{cases} \tag{2.5}$$

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

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0, \\ 0 & x \leq 0. \end{cases} \tag{2.6}$$



lambda=1
lambda=0.5
lambda=3
f(x)
x
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10

in R

```
dexp, rexp, pexp, qexp
```

↑Code ↓Code

从图 (2.5) 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.

设 X 表示某种电子元件的寿命，$F(x)$ 为其分布函数。若假设元件无老化，即元件在时刻 x 正常工作的条件下，其失效率保持为某个常数 λ，与 x 无关。试证明 X 服从指数分布。

↑Example

↓Example

解：失效率即单位时间内失效的概率，因此由题设知

$$P(x \leq X \leq x+h|X>x)/h = \lambda, \quad h \to 0$$

因为

$$P(x \leq X \leq x+h|X>x) = \frac{P(\{x \leq X \leq x+h\}\{X>x\})}{P(X>x)} = \frac{F(x+h)-F(x)}{1-F(x)}$$

所以有

$$\lim_{h \to 0} P(x \leq X \leq x+h|X>x)/h = \frac{F'(x)}{1-F(x)} = \lambda$$

即得到微分方程 $\frac{F'(x)}{1-F(x)} = \lambda$, 解此方程得到

$$F(x) = 1 - e^{-\lambda x}$$

从而结论得证。

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

$$P(X > s + t \mid X > s) = P(X > t). \tag{2.7}$$

即寿命是无老化的. 可以证明，指数分布是唯一具有性质 (2.7) 的连续型分布.

2.3.3 均匀分布

设 $a < b$，如果分布 $F(x)$ 具有密度函数

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b\ , \\ 0 & \text{其它}, \end{cases} \tag{2.8}$$

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

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

in R

```
dunif, runif, punif, qunif
```

↑Code
↓Code

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.

第二讲：随机变量及其分布

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第二章随机变量及其分布

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．

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.

↑Example

↓Example

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.

↑Example

↓Example

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.

Definition

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**、**连续型**以及**其他类型**.

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

Definition

$$p(j_1,\cdots,j_n)=P(X_1=a_{1j_1},\ldots,X_n=a_{nj_n}),\ j_1,...,j_n=1,2,... \tag{2.1}$$

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

↑Example

设 $A_1,\cdots,A_n$ 为某一实验下的完备事件群，即 $A_1,\cdots,A_n$ 两两互斥且和为 Ω。记 $p_k=P(A_k)(k=1,\dots,n)$，则 $p_k\ge 0, p_1+\cdots+p_n=1$。现将实验独立的重复作 N 次，分别用 X_i 表示事件 A_i 出现的次数 $(i=1,\cdots,n)$。则 $X=(X_1,\dots,X_n)$ 为一离散型随机向量，试求 X 的概率函数。此分布律称为多项分布，记为 $M(N;p_1,\dots,p_n)$.

↓Example

解: 由于试验独立进行, 总的结果数为 N，记结果 A_i 出现的次数为 k_i，则 $k_1+\cdots+k_n=N$。因此相当于多组组合，所以

$$
\begin{aligned}
P(X_1=k_1,\cdots,X_n=k_n) &= \frac{N!}{k_1!\cdots k_n!}P(\underline{A_1\cdots A_1}\dots\underline{A_n\cdots A_n}) \\
&= \frac{N!}{k_1!\cdots k_n!}p_1^{k_1}\cdots p_n^{k_n},
\end{aligned}
$$

其中 $k_1,\dots,k_n$ 为非负整数且 $k_1+\cdots+k_n=N$.

我们来看一下 X_i 的分布：此时我们把试验结果分为两类, A_i 和 $\bar{A}_i$，则显然就是一个 N 重贝努里试验，因此

$$P(X_i = k_i) = \binom{N}{k_i} p_i^{k_i} (1 - p_i)^{N-k_i}, \quad k_i = 1, \cdots, N.$$

类似我们也可以找出 $(X_i, X_j)(i \neq j)$ 的联合分布律，即为 $M(N, p_i, p_j, 1-p_i - p_j)$.

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) 的概率函数可以用下表表示:

$Y \backslash X$	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

↑Example

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

$$p(x,y)=\frac{\binom{6}{x}\binom{7}{y}\binom{5}{4-x-y}}{\binom{18}{4}},\ 0\leq x+y\leq 4. \tag{2.2}$$

↓Example

以列联表表示, 即为

Y \ 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.

Definition

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

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

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.

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

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \tag{2.3}$$

为 n 维随机变量 X 的 (联合) 分布函数.

Definition

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 $1 \leq j \leq n$ 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,

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

对高维离散型随机变量, 一般我们不使用分布函数.

↑Example

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.

↓Example

↑Example

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

↓Example

2.5 边缘分布

设 $(X_1, ..., X_n) \sim F$ 已知. 令 $(i_1, ..., i_m) \subset (1, ..., n)$, 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. 如何得到该分布?

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

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

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, 我们称其为 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.

n 维场合:

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 $i_1 < \cdots < i_m$ 为 $1, ..., n$ 的任一子集, 则 $X_{i_1}, ..., X_{i_m}$ 的概率函数为

$$
\begin{aligned}
& p_{i_1 \ldots i_m}(j_{i_1}, ..., j_{i_m}) = P(X_{i_1} = a_{i_1 j_{i_1}}, ..., X_{i_m} = a_{i_m j_{i_m}}) \\
= \ & P(X_{i_1} = a_{i_1 j_{i_1}}, ..., X_{i_m} = a_{i_m j_{i_m}}) \\
= \ & \sum_{j_{i_{m+1}}, ..., j_{i_n}} P(X_1 = a_{1 j_1}, ..., X_{i_1} = a_{i_1 j_{i_1}}, ..., X_{i_m} = a_{i_m j_{i_m}}, \\
& X_{i_{m+1}} = a_{i_{m+1} j_{i_{m+1}}}, ..., X_n = a_{n j_n}) \\
= \ & \sum_{\text{除} j_{i_1}, ..., j_{i_m} \text{的所有}} p(j_1, ..., j_n).
\end{aligned}
$$

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

袋中有 5 张外形相同的卡片，其中 3 张写上数字"0"，另 2 张写上"1"。现从袋中任取两张卡片，分别以 ξ, η 表示第一张和第二张卡片上的数字，试求分别在有放回和不放回两种情形下 (ξ, η) 的联合分布律及边际分布律.

↑Example ↓Example

解：简单计算得到

$\eta\backslash\xi$	0	1	$p_{\cdot j}$
0	$9/25$	$6/25$	$3/5$
1	$6/25$	$4/25$	$2/5$
$p_{i\cdot}$	$3/5$	$2/5$	1

$\eta\backslash\xi$	0	1	$p_{\cdot j}$
0	$6/20$	$6/20$	$3/5$
1	$6/20$	$2/20$	$2/5$
$p_{i\cdot}$	$3/5$	$2/5$	1

这个例子说明**边际分布律不能决定联合分布律**。

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

$$
\begin{aligned}
P(x_1 \le X \le x_2) &= P(x_1 \le X \le x_2, -\infty < Y < +\infty) \\
&= \int_{-\infty}^{+\infty} \int_{x_1}^{x_2} f(u,v)dudv \\
&= \int_{x_1}^{x_2} f_X(u)du, \qquad (2.4)
\end{aligned}
$$

其中

$$
f_X(u) = \int_{-\infty}^{+\infty} f(u,v)dv. \qquad (2.5)
$$

从 (2.4) 我们可以看出, X 的边缘密度函数即为 (2.5). 类似地, Y 的边缘密度函数为

$$
f_Y(u) = \int_{-\infty}^{+\infty} f(u,v)du. \qquad (2.6)
$$

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

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

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

↑Example

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 $N(b,\sigma_2^2)$.

↓Example

例2.5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.

↑Example

考虑两个概率密度函数

$$\begin{aligned} p(x,y) &= x+y, \qquad 0<x,y<1 \\ q(x,y) &= (x+\frac{1}{2})(y+\frac{1}{2}), \qquad 0<x,y<1 \end{aligned}$$

试求边际概率密度。

↓Example

解：易得所求边际概率密度都是如下形式

$$f(t)=t+\frac{1}{2}, \quad 0<t<1.$$

说明**边际概率密度不能决定联合概率密度**。

↑Example

设 (X,Y) 的联合概率密度有形式 $(\forall(x,y)\in\mathbb{R}^2)$

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

其中 $-\infty<a,b<\infty; 0<\sigma_1,\sigma_2<\infty; -1\le\rho\le 1$. 则称 (X,Y) 服从参数为 $a,b,\sigma_1,\sigma_2,\rho$ 的二元正态分布，记为 $N(a,b,\sigma_1^2,\sigma_2^2,\rho)$. 试计算 X 和 Y 的边际概率密度。

↓Example

解:

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a)^2}{\sigma_1^2}\right.\right. \\
&\quad \left.\left. -2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} exp\left\{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right\} dv \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1} exp\{-\frac{1}{2}[(\frac{v-\rho u}{\sqrt{1-\rho^2}})^2+u^2]\}dv \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} exp\{-\frac{(x-a)^2}{2\sigma_1^2}\}
\end{aligned}
$$

即 $X \sim N(a,\sigma_1^2)$. 类似可得 $Y \sim N(b,\sigma_2^2)$, 其边际概率密度为 $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} exp\{-\frac{(y-b)^2}{2\sigma_2^2}\}$.

第二讲：随机变量及其分布

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第二章随机变量及其分布

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 $Y_i = g_i(X_1, X_2, \cdots, X_n)$，$i = 1, 2, \cdots, m$。

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

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

↑Example

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。

↓Example

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**

↑Example

设 $X \sim B(n,p)$，$Y \sim B(m,p)$ 且 X 和 Y 相互独立，则 $X+Y \sim B(n+m,p)$。

↓Example

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。

↑Example

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。

↓Example

2. 连续型随机变量的情形

定理 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* ∞*)*, 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

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

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

↑Example
↓Example

解: 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。

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

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:

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

设 $X \sim N(0,1)$，求 $Y = X^2$ 的概率密度。

↑Example
↓Example

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. 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 (ξ_1, ξ_2) 与 (ζ_1, ζ_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 (ζ_1, ζ_2) 亦为连续型随机向量，且其联合概率密度为

$$q(y_1, y_2) = \begin{cases} p\left(h_1(y_1, y_2), h_n(y_1, y_2)\right)|J|, & (y_1, y_2) \in \mathbb{D}, \\ 0, & (y_1, y_2) \notin \mathbb{D}, \end{cases} \tag{2.1}$$

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 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

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

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

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

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.

↑Example

↓Example

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 (2.2) 式得知, (ρ,θ) 的联合密度为

$$q(r,t)=\frac{1}{2\pi}r\exp\left\{-\frac{r^2}{2}\right\}=q_1(r)q_2(t),\quad r>0,\quad t\in[0,2\pi). \quad (2.3)$$

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 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

$$F(z) = P(X+Y \le z) = \iint_{x+y\le z} f(x,y)dxdy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} f(x,y)dy$$

$$= \int_{-\infty}^{\infty} du \int_{-\infty}^{z} f(x, t-x)dt = \int_{-\infty}^{z} \left\{ \int_{-\infty}^{\infty} f(x,t-x)dx \right\} dt.$$

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 (2.1) 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

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

称此公式为**卷积公式**。

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 $P(X \le Y)$。

$\overline{\uparrow}$Example

$\underline{\downarrow}$Example

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.

↑Example
↓Example

设 $X_1,\ldots,X_n i.i.d \sim N(0,1)$, 试求 $Y_n=\sum_{i=1}^n X_i^2$ 的分布.

解: 由前例知 $Y_1=X_1^2$ 的概率密度函数为

$$f_1(y)=\frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2}I(y>0)=\boxed{\frac{1}{2^{1/2}\Gamma(\frac{1}{2})}y^{1/2-1}e^{-y/2}I(y>0)}$$

由卷积公式知 $Y_2=X_1^2+X_2^2$ 的概率密度函数为

$$\begin{aligned}f_2(z)&=\int_R f_1(y)f_2(z-y)dy\\&=\frac{1}{2\Gamma^2(\frac{1}{2})}e^{-z/2}I(z>0)\int_0^1 t^{-1/2}(1-t)^{-1/2}dt\\&=\boxed{\frac{1}{2^{2/2}\Gamma(\frac{2}{2})}z^{2/2-1}e^{-z/2}I(z>0)}\end{aligned}$$

从而由归纳法, 假设 $Y_{n-1}=X_1^2+\cdots+X_{n-1}^2$ 的概率密度函数为

$$f_{n-1}(z) = \boxed{\frac{1}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})} z^{(n-1)/2-1} e^{-z/2} I(z>0)}$$

则 $Y_n = Y_{n-1} + X_n^2$ 的概率密度函数可由卷积公式得

$$\begin{aligned} f_n(z) &= \int_R f_{n-1}(y) f_1(z-y) dy \\ &= \frac{1}{2^{n/2}\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})} z^{n/2-1} e^{-z/2} I(z>0) \int_0^1 t^{(n-1)/2-1}(1-t)^{1/2-1} dt \\ &= \boxed{\frac{1}{2^{n/2}\Gamma(\frac{n}{2})} z^{n/2-1} e^{-z/2} I(z>0)} \end{aligned}$$

由归纳法得 Y_n 的密度函数. 称 Y_n 的分布为自由度 n 的卡方分布, 记为 $Y_n \sim \chi_n^2$.

- χ_n^2 具有再生性
- $X \sim \chi_n^2$, 则 $EX = n, Var(X) = 2n$.
- http://en.wikipedia.org/wiki/Chi-squared_distribution

一些连续型随机变量，也有再生性性质。

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 $X = \sum_{i=1}^{n} a_i X_i + b$, 则有: $X \sim N(\mu, \sigma^2)$, 其中 $\mu = \sum_{i=1}^{n} a_i \mu_i + b$, $\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$.

↑Example

↓Example

设 $X_1 \sim \chi_n^2, X_2 \sim \chi_m^2$, 且 X_1 和 X_2 相互独立，则 $X_1 + X_2 \sim \chi_{n+m}^2$.

↑Example

↓Example

我们把具有再生性性质的分布总结一下为

- 二项分布 (关于试验次数具有再生性)
- $Poisson$ 分布 (关于参数 λ 具有再生性)
- $Pascal$ 分布 (关于成功次数 r 具有再生性)
- 正态分布 (关于两个参数都具有再生性)
- χ^2 分布具有再生性

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 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

$$
\begin{aligned}
p_{\frac{\xi}{\eta}}(x) &= \int_{-\infty}^{\infty} |t| f(xt, t)dt, \quad \forall\ x \in \mathbb{R}. \\
\text{而 } p_{\frac{\eta}{\xi}}(x) &= \int_{-\infty}^{\infty} |u| f(u, xu)du, \quad \forall\ x \in \mathbb{R}.
\end{aligned}
\tag{2.4}
$$

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.

↑Example

↓Example

解: 我们利用 (2.4) 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 (2.4) 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。

设 $X_1 \sim N(0,1)$，$X_2 \sim \chi_n^2$，且 X_1 与 X_2 独立。求 $Y = \frac{X_1}{\sqrt{X_2/n}}$ 的概率密度函数. (记 $Y \sim t_n$, 称为自由度为 n 的 t 分布)。

↑Example
↓Example

解: 先求 $Z = \sqrt{X_2/n}$ 的密度 $g(z)$:

$$g(z) = 2nzf_{X_2}(nz^2)I(z > 0)$$

其中 f_{X_2} 为 X_2 的密度函数. 利用商的密度变换公式, 可得

$$\begin{aligned}
f_Y(y) = f_{X_1/Z}(y) &= \int_R |t|\phi(yt)g(t)dt \\
&= \int_R |t| \frac{1}{\sqrt{2\pi}} e^{-(yt)^2/2} \frac{2nt}{2^{n/2}\Gamma(\frac{n}{2})} (nt^2)^{n/2-1} e^{-nt^2/2} I(t > 0)dt \\
&= \frac{1}{\sqrt{2\pi}} \frac{2n^{n/2}}{2^{n/2}\Gamma(\frac{n}{2})} \int_0^\infty t^n e^{-(y^2+n)t^2/2} dt
\end{aligned}$$

令 $x=(n+y^2)t^2/2$, 则上述积分为

$$f_Y(y)=\frac{1}{\sqrt{2\pi}}\frac{2n^{n/2}}{2^{n/2}\Gamma(\frac{n}{2})}\int_0^\infty x^{(n-1)/2}e^{-x}dx\cdot\frac{1}{2}\left(\frac{2}{n+y^2}\right)^{(n+1)/2}$$
$$=\boxed{\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{y^2}{n}\right)^{-\frac{n+1}{2}}}$$

- t_n 的分布关于原点对称
- $lim_{n\to\infty}f_Y(y)=\phi(y)$.
- http://en.wikipedia.org/wiki/Student's_t-distribution

设 $X_1 \sim \chi_n^2$, $X_2 \sim \chi_m^2$, 且 X_1 与 X_2 独立, 求 $Y = \frac{X_1/n}{X_2/m}$ 的概率密度函数. (记 $Y \sim F_{n,m}$, 称为自由度为 n, m 的 F 分布)。

↑Example

↓Example

解: 由密度变换公式易知 X_1/n 的密度函数为 $ng_n(nz)$, 类似 X_2/m 的密度为 $mg_m(mz)$, 其中 g_n 为自由度 n 的 χ_n^2 的密度函数. 从而 Y 的密度函数为

$$
\begin{aligned}
f_Y(y) &= \int_R |t| n g_n(nty) m g_m(mt) I(t>0) dz \\
&= [2^{(n+m)/2}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})]^{-1} n^{n/2} m^{m/2} y^{n/2-1} \\
&\quad \cdot \int_0^\infty e^{-(ny+m)t/2} t^{(n+m)/2-1} dt \\
&= \boxed{\frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} n^{n/2} m^{m/2} y^{n/2-1} (ny+m)^{-(n+m)/2}}
\end{aligned}
$$

- http://en.wikipedia.org/wiki/F-distribution

极小值和极大值的分布

对于 n 个随机变量 $X_1, ..., X_n$, 我们可以考察它们的最大值和最小值:

$$X_{(n)} = \max\{X_1, ..., X_n\},$$

$$X_{(1)} = \min\{X_1, ..., X_n\}.$$

如此定义的 $X_{(n)}$ 与 $X_{(1)}$ 也是随机变量.

当 $X_1, ..., X_n$ 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 $X_{(n)}$ 与 $X_{(1)}$ 的分布函数 $F_{X_{(n)}}(x)$ 和 $F_{X_{(1)}}(x)$.

事实上,

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \le x) = P(\max\{X_1, \cdots, X_n\} \le x) = P\left(\bigcap_{k=1}^{n} (X_k \le x)\right) \\ &= \prod_{k=1}^{n} P(X_k \le x) = \prod_{k=1}^{n} F_k(x); \end{aligned} \tag{2.5}$$

而利用关系式

$$(X_{(1)} > x) = (X_1 > x,\ \dots, X_n > x) = \bigcap_{k=1}^{n} (X_k > x)$$

可得

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \le x) = 1 - P(X_{(1)} > x) = 1 - P\left(\bigcap_{k=1}^{n} (X_k > x)\right) \\ &= 1 - \prod_{k=1}^{n} P(X_k > x) = 1 - \prod_{k=1}^{n} (1 - F_k(x)). \end{aligned} \tag{2.6}$$

在 $X_1, \ldots, X_n$ 还是同分布时候, (2.5) 和 (2.6) 还可以简化.

设 $X_1, \ldots, X_n i.i.d \sim U(0,\theta), \theta > 0$, 求 $X_{(n)} = max_{1\le i\le n} X_i$ 的密度函数.

↑Example

↓Example

解: 由于对任意 $0 < x < \theta$,

$$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(\max\{X_1, \cdots, X_n\} \le x) = \prod_{k=1}^{n} F_k(x)$$

$$= \left(\frac{x}{\theta}\right)^n$$

从而所求密度为

$$g(x) = F'_{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n} I(0 < x < \theta).$$

目前我们接触到的分布的关系为

- n 个独立同分布 $B(1,p)$ 的 0-1 分布随机变量之和为二项分布 $B(n,p)$;
- 有限个独立二项随机变量 (成功的概率相同) 之和仍为二项分布;
- 有限个独立的 $Poisson$ 分布随机变量之和服从 $Poisson$ 分布，参数相加;
- r 个独立同分布几何分布 $G(p)$ 的随机变量之和服从参数为 r 和 p 的 $Pascal$ 分布;
- 任意有限个独立的正态分布随机变量的线性组合仍然服从正态分布;

第三讲：随机变量的数字特征

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第三章随机变量的数字特征

随机变量的性质描述

1. 随机变量的**分布函数是对随机变量的概率性质最完整的刻画.**

2. 有些时候我们更关心随机变量的某方面 “特征”(完全由分布函数决定的):

    - 某行业工人的平均工资 (这里工资的分布情况不是最关心的), 或者某行业工人的工资散布程度

3. **能够刻画随机变量某些方面的性质特征的量称为随机变量的数字特征。**

    - 度量” 中心”: **期望, 中位数**
    - 度量散布程度: **方差, 绝对偏差, 极差**
    - 分布形状: **偏度系数, 峰度系数**
    - 相关程度: **相关系数**

3.1 数学期望 (均值) 及中位数

3.1.1 数学期望 (Expectation)

数学期望也称均值 (Mean), **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

↑Example

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?

↓Example

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

$$X\text{的可能值与其概率之积的累加}$$

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.

↑Example

甲乙两人射击水平如下所示

甲:

击中环数	8	9	10
概率	0.3	0.1	0.6

乙:

击中环数	8	9	10
概率	0.2	0.5	0.3

试问两人谁的水平高?

↓Example

假设两人分别射击 N 次, 则他们各自射击的总环数大概为

$$\text{甲: } 8 * 0.3N + 9 * 0.1N + 10 * 0.6N = 9.3N$$

$$\text{乙: } 8 * 0.2N + 9 * 0.5N + 10 * 0.3N = 9.1N$$

因此, 在 N 次射击后, 两人的平均击中环数分别为 9.3 和 9.1, 因此甲的水平稍高一些.

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

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

Definition

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

为随机变量 X 的数学期望（均值），用符号 EX 表示．若 $\sum_{i=1}^{\infty} |x_i| p_i = +\infty$，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
不妨设连续型随机变量 X 的密度 $f(x)$ 的非零取值范围为

$(a,b), a<b$ 可以为 $\mp\infty$, 则可以通过将 X 离散化来考虑 X 的期望:

1. 取点集 $\{x_i\}$, 使得 $a = x_0 < x_1 < \ldots < x_n = b$, 区间长为 $\Delta x_i = x_i - x_{i-1}$.

2. 定义一个新的离散型随机变量 X', 其所有可能取值点为 $\{t_i\}$, $x_{i-1} < t_i \leq x_i$ 且有分布律

$$P(X' = t_i) = p_i = P(x_{i-1} < X \leq x_i) \approx f(t_i)\Delta x_i$$

3. 从而有离散型随机变量期望的定义有:$(\Delta x_i \to 0)$

$$EX' = \sum t_i p_i \approx \sum t_i f(t_i)\Delta x_i \to \int_R x f(x)dx := EX < \infty,$$

$$\text{如果} \sum |x_i| p_i \approx \sum |t_i| f(t_i)\Delta x_i \to \int_R |x| f(x)dx < \infty.$$

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．

Definition

下面求解几种常见分布的数学期望.

1. 二项分布 $X \sim B(n,p)$:

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

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 $X \sim U[a, b]$:

$$EX = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

5. 指数分布 $X \sim Exp(\lambda)$:

$$EX = \int_0^\infty x\lambda e^{-\lambda x} dx = 1/\lambda.$$

6. 卡方分布 $X \sim \chi_n^2$:

$$EX = n.$$

7. t 分布 $X \sim t_n$:

$$EX = 0.$$

设 $r.v.\ X$ 的分布律为

$$P\left(X=(-1)^k\frac{2^k}{k}\right)=\frac{1}{2^k},\quad k=1,2,\cdots$$

则 X 的数学期望不存在。

↑Example

↓Example

解: 由于

$$\sum_{k=1}^{\infty}|(-1)^k\frac{2^k}{k}|\frac{1}{2^k}=\sum_{k=1}^{\infty}\frac{1}{k}=+\infty$$

因此 X 的数学期望不存在。而尽管

$$\sum_{k=1}^{\infty}(-1)^k\frac{2^k}{k}\frac{1}{2^k}=\sum_{k=1}^{\infty}(-1)^k\frac{1}{k}=-ln2.$$

↑Example

(Cauchy 分布) 设

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathcal{R},$$

则: 该分布的期望不存在.

↓Example

解: 容易看出,$p(x)$ 非负, 并且

$$\int_{-\infty}^{\infty} p(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+x^2}dx = \frac{1}{\pi}\arctan x\,|_{-\infty}^{\infty} = 1,$$

所以 $p(x)$ 是一个密度函数 (称为 Cauchy 分布), 但是

$$\int_{-\infty}^{\infty} |x|p(x)dx = \frac{2}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2}dx = \infty,$$

所以 Cauchy 分布的期望不存在. #

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.

假设随机变量 $X \sim B(n,p)$, 求 EX.

↑Example
↓Example

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

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

假设 c 为常数, 则 $EcX = cEX$.

↑Example
↓Example

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

↑Example
↓Example

解: 由 $X \sim N(0,1)$, 我们有

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

所以, $EY = EX^2 + 1 = 2$.

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.

↑Example ↓Example

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 (Conditional Mean)

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 $E(Y|x)$.

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

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

Definition

期望所具有的性质条件期望同样满足.

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

↑Example
↓Example

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

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

[**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.

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.

↑Example

↓Example

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 $EX = 15$.

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

↑Example
↓Example

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 (Median)

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.

称 m 为连续型随机变量 X 的中位数, 如果

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

Definition

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.

- 中位数总是存在的.
- 和期望值相比中位数的一个优点是它受个别特别大或特别小的值的影响很小, 而期望则不然: 收入差距非常大时, 中位数比均值更加有效

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.

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

↑Example
↓Example

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.

中位数的定义是 p 分位数定义的特例:

设 $0 < p < 1$, 称 μ_p 是随机变量 X 的 p 分位数, 如果

$$P(X \le \mu_p) \ge p, \quad P(X \ge \mu_p) \ge 1 - p.$$

Definition

第三讲：随机变量的数字特征

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第三章随机变量的数字特征

3.2 方差, 相关系数以及其他数字特征

3.2.1 方差 (Variance)

方差是刻画随机变量**在其中心位置附近的散布程度**. 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.**

设 X 为随机变量, 分布为 F, 若 X 平方可积, 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**.

Definition

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 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 $P(X = c) = 1$, 其中 $c = EX$. 此时, 我们称 X 退化到常数 c.
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 $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$.

证明上述定理，我们介绍一个引理。

引理 1. *如果 ξ 为退化于 0 的随机变量，则有 $E\xi^2 = 0$；反之，如果随机变量 ξ 的 2 阶矩存在而且 $E\xi^2 = 0$，则 ξ 必为退化于 0 的随机变量.*

Proof. 如果 ξ 为退化于 0 的随机变量，则有 $P(\xi = 0) = 1$, 故有 $E\xi^2 = 0$。反之，如果随机变量 ξ 平方可积，并且 $E\xi^2 = 0$，但是 ξ 不退化于 0，则有 $P(\xi = 0) < 1$。那么就存在 $\delta > 0$ 和 $0 < \epsilon < 1$，使得 $P(|\xi| > \delta) > \epsilon$，于是 $E\xi^2 > \delta^2\epsilon$。导致矛盾，所以 ξ 必退化到 0. □

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 $X \sim N(\mu,\sigma^2)$:

$$VarX = \sigma^2$$

我们称

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

为 X 的**标准化随机变量**. 易见 $EX^* = 0, Var(X^*) = 1$.

Definition

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 (Moments) 的概念，并将之与我们前面所说的期望、方差建立联系.

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.

Definition

比较重要的有两个情况:

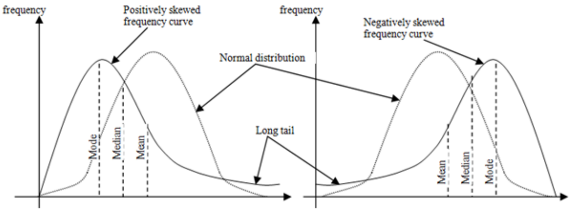
1. $c=0$. 这时 $\alpha_k = EX^r$ 称为 X 的 r 阶**原点矩**.

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 $Var(X)$.

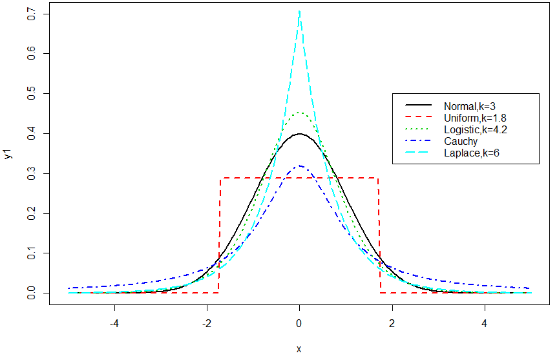
- **偏度系数**

$$\gamma_1 = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{\mathrm{E}\left[(X-\mu)^3\right]}{\left(\mathrm{E}\left[(X-\mu)^2\right]\right)^{3/2}}$$



- **峰度系数**

$$\gamma_2 = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4} = \frac{\mathrm{E}\left[(X-\mu)^4\right]}{\left(\mathrm{E}\left[(X-\mu)^2\right]\right)^2}$$



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.

注意到

$$Var(X+Y)=Var(X)+Var(Y)+2E(X-EX)(Y-EY)$$

即

$X+Y$ 的波动性 $=X$ 的波动性 $+Y$ 的波动性 $+X$ 和 Y 的相关性

3.3.1 协方差

如果随机变量 X 和 Y 平方可积，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.

Definition

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

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

如果 $\xi_1,\cdots,\xi_n$ 是定义在同一概率空间下的随机变量，并且其中每个随机变量都是平方可积的。称矩阵

$$\begin{aligned}\Sigma &= (b_{ij})=(cov(\xi_i,\xi_j))\\ &= \begin{pmatrix} D(\xi_1) & cov(\xi_1,\xi_2) & \cdots & cov(\xi_1,\xi_n)\\ cov(\xi_2,\xi_1) & D(\xi_2) & \cdots & cov(\xi_2,\xi_n)\\ \vdots & \vdots & \ddots & \vdots\\ cov(\xi_n,\xi_1) & cov(\xi_n,\xi_2) & \cdots & D(\xi_n)\end{pmatrix}\end{aligned}$$

为 $\xi_1,\cdots,\xi_n$ 的协方差矩阵。显然 $\Sigma\geq 0$。

设 $(X,Y) \sim N(a,b,\sigma_1^2,\sigma_2^2,\rho)$，则 (X,Y) 的协方差矩阵为

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

↑Example

↓Example

3.3.2 相关系数

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 (Coefficient of correlation).
当 $\rho_{X,Y} = 0$ 时，则称 X 与 Y 不相关 (uncorrelated).

Definition

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.

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

↑Example
↓Example

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

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

[**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.

为证明 2, 我们看如下引理。

引理 2. [*Cauchy − Schwarz Inequality*] 设 ξ, η 均平方可积, 则有

$$[E\xi\eta]^2 \leq E\xi^2 E\eta^2$$

等号成立当且仅当 $P(\xi = t_0\eta) = 1$, 其中 t_0 为一常数。

Proof. 易知, 对任何 $t \in \mathcal{R}$, 都有

$$g(t) := E\eta^2 \cdot t^2 - 2E\xi\eta \cdot t + E\xi^2 = E(\xi - t\eta)^2 \geq 0 \ ,$$

所以二次函数 $g(t)$ 的判别式

$$\Delta = 4(E\xi\eta)^2 - 4E\xi^2 \cdot E\eta^2 \leq 0,$$

故得不等式.

如果存在 $t_0 \in \mathcal{R}$, 使得 $P(\xi = t_0\eta) = 1$, 显然就有

$$(E\xi\eta)^2 = E\xi^2 E\eta^2.$$

反之, 如果不等式等号成立, 那么方程 $g(t) = 0$ 有唯一的实根 t_0, 即有

$$E(\xi - t_0\eta)^2 = g(t_0) = 0,$$

于是由引理1知 $\xi - t_0\eta$ 是退化于 0 的随机变量, 即有 $P(\xi = t_0\eta) = 1$. □

推论 1. *设随机变量 ξ, η 平方可积, 则有*

$$cov(\xi, \eta) \leq \sqrt{D\xi} \cdot \sqrt{D\eta},$$

并且等号成立, 当且仅当存在 $t_0 \in \mathcal{R}$, 使得 $P(\xi = t_0\eta) = 1$.

设 $X \sim U(-\frac{1}{2},\frac{1}{2})$, 而 $Y = cosX$, 证明 X, Y 不相关.
但是 X, Y 之间存在着非线性的函数关系.

↑Example

↓Example

证:

由于 $EX = 0$,

$$E(XY) = E(Xcos(X)) = \int_{-1/2}^{1/2} xcosxdx = 0$$

所以

$$Cov(X,Y) = EXY - EXEY = 0$$

即 X, Y 不相关. 但是 X, Y 之间存在着非线性的函数关系.

定理 2. *对任何非退化的随机变量 ξ, η 平方可积, 如下四个命题相互等价:*

(1) ξ 与 η 不相关; (2) $cov(\xi,\eta) = 0$;
(3) $E\xi\eta = E\xi E\eta$; (4) $Var(\xi+\eta) = Var(\xi) + Var(\eta)$.

下面我们来讨论不相关与独立性之间的关系.

定理 3. 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.

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.

↑Example

↓Example

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

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

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.

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.

↑Example

↓Example

[**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 $E|X - EX|$
- 矩母函数 $g(t) = Ee^{tX}$, 其中 $t \in \mathbb{R}$.
- 特征函数 $\phi(t) = Ee^{itX}$, 其中 $t \in \mathbb{R}$, i 为虚数.

定理 4. 对任何随机变量 X, Y, 分别有分布函数 F_X, F_Y 和特征函数 ϕ_X, ϕ_Y, 则

$$F_X = F_Y \leftrightarrow \phi_X = \phi_Y$$

第四讲: 大数定律和中心极限定理

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第四章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. **大数定律**, 是概率论中讨论随机变量和的平均值的收敛情况, 是数理统计学中参数估计的理论基础. **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.

4.1 大数定律

如果对任何 $\varepsilon > 0$, 都有

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

Definition

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 ξ, 记为 $\xi_n \xrightarrow{p} \xi$.

定理 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

如果以 ζ_n 表示 n 重 Bernoulli 试验中的成功次数, 则有

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

↑Example ↓Example

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1, 我们需要如下的 Chebyshev 不等式:

引理 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.

4.2 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.

定理 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

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

Proof. 由于标准正态分布的特征函数为 $f(t)=e^{-t^2/2}$，因此我们只需证明 $\eta_n=\sum\limits_{i=1}^{n}\frac{X_i-\mu}{\sigma}$ 的特征函数的极限是 $f(t)$ 就可以了。

记 $\{X_i-\mu\}$ 的共同特征函数为 $g(t)$，则

$$g\Big(\frac{t}{\sigma\sqrt{n}}\Big)=1-\frac{t^2}{2n}+o\Big(\frac{t^2}{n}\Big)$$

而 η_n 的特征函数为 $g^n(\frac{t}{\sigma\sqrt{n}})$. 由于

$$\Big|g^n\Big(\frac{t}{\sigma\sqrt{n}}\Big)-\Big(1-\frac{t^2}{2n}\Big)^n\Big|\leq n\Big|g\Big(\frac{t}{\sigma\sqrt{n}}\Big)-\Big(1-\frac{t^2}{2n}\Big)\Big|=no\Big(\frac{t^2}{n}\Big)\longrightarrow 0$$

所以

$$\lim_{n\to\infty}g^n\Big(\frac{t}{\sigma\sqrt{n}}\Big)=e^{-t^2/2}$$

即

$$\lim_{n\to\infty}P(\eta_n\leq x)=\Phi(x)$$

□

定理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2, 我们很容易得到如下推论

定理 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

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

定理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2, 近似地有

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

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

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

设一考生参加 100 道题的英语标准化考试 (每道题均为有两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 50 的概率.

↑Example

↓Example

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

$$P(X_1 = 0) = P(X_1 = 1) = 0.5.$$

利用中心极限定理, 有

$$\begin{aligned} P(X_1 + \cdots + X_{100} &\geq 50) = P\left(\frac{X_1 + \cdots + X_{100} - 100 * 0.5}{\sqrt{100 * 0.5 * 0.5}} \geq 0\right) \\ &= 1 - \Phi(0) = 1/2. \end{aligned}$$

↑Example

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

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

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.

↓Example

求极限 $\lim\limits_{n\to\infty}\sum\limits_{k=1}^{n}\frac{n^k}{k!}e^{-n}$.

↑Example
↓Example

定理 4. 设 $X\sim B(n,p)$，则有

$$\lim_{n\to\infty}P(\frac{X-np}{\sqrt{npq}}\le x)=\Phi(x),\quad \forall\ x\in\mathbb{R}$$

即

$$\frac{X-np}{\sqrt{npq}}\overset{asy.}{\sim}N(0,1).$$

Proof. 由二项分布随机变量和 0-1 分布随机变量之间的关系及中心极限定理易证。 □

在仅有独立性和二阶矩有限场合下，我们有

定理 5. 设 $\{X_n\}$ 为独立的随机变量序列，而且具有数学期望 $EX_k = \mu_k$ 和方差 $D(X_k) = \sigma_k^2 < \infty$, $k = 1, 2, \cdots$. 记

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2$$

若存在正数 δ，使得当 $n \to \infty$ 时

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|X_k - EX_k|^{2+\delta} \to 0$$

则有

$$\lim_{n\to\infty} P\left(\sum_{k=1}^{n} \frac{X_k - \mu_k}{B_n} \leq x\right) = \Phi(x) \qquad \forall x \in \mathbb{R} \qquad (4.1)$$

例题参考课本.

如果独立随机变量序列 $\{X_n,\ n\in\mathbb{N}\}$ 同上述定理，并且对任何 $\tau>0$，都有

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}E\left\{(X_k-a_k)^2I(|X_k-a_k|\geq\tau B_n)\right\}=0, \tag{4.2}$$

则称该随机变量序列满足 `Linderberg` 条件.

Definition

定理 6. 设随机变量序列 $\{X_n\}$ 满足 *Linderberg* 条件 *(4.2)*, 则 $\{X_n\}$ 满足中心极限定理, 即 *(4.1)* 式成立.

第五章：数理统计的基本概念与抽样分布

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第五章：数理统计的基本概念与抽样分布

4.1　引言

4.1.1　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.

↑Example

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.

↓Example

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》.

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.

↑Example

↓Example

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:

↑Example

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 \ 水平	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.

↓Example

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.

↑Example

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.

↓Example

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.

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?

↑Example

↓Example

(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.

4.2 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.

↑Example

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**.

↓Example

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:

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

由于每个个体在抽样时的出现是随机的, 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:

Definition

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

$$X_1, \cdots, X_n \text{ i.i.d. } \sim F \tag{4.1}$$

若 F 有密度 f, 可记为

$$X_1, \cdots, X_n \ \text{ i.i.d. } \sim f \tag{4.2}$$

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

$$X_1, \cdots, X_n \text{ i.i.d. } \sim X \tag{4.3}$$

(4.1)、(4.1) 和 (4.3) 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. 样本空间

我们知道样本是由总体中抽取的一部分个体组成. 设 $X = (X_1, \cdots, X_n)$ 是从总体中抽取的一个样本, 其样本空间如下:

样本 $X = (X_1, \cdots, X_n)$ 可能取值的全体成为**样本空间**, 记为 $\mathscr{X}$.

Definition

例如在前面称重例中, 样本空间为 $\mathscr{X} = \{(x_1, \cdots, x_5) : 0 < x_i < \infty,\ i = 1, 2, \cdots 5\}$, 也可以写成 $\mathscr{X} = \{(x_1, \cdots, x_5) : -\infty < x_i < \infty,\ i = 1, 2, \cdots 5\}$. 虽然物重不可以取负数, 但这无关紧要, 因为在考虑样本分布时, 可令样本取负值的概率为 0. 再看下例:

↑Example

打靶试验，每次打三发，考察中靶的环数．如样本 $X=(5,1,9)$ 表示三次打靶分别中 5 环、1 环和 9 环．此时样本空间为

$$\mathscr{X}=\{(x_1,x_2,x_3):\ x_i=0,1,2,\cdots 10,\ i=1,2,3\}$$

这个样本空间中样本点数是有限的，上例称重问题中样本空间中的样本点数是无限的．

↓Example

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 $0\le X_i\le 10$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:

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.

Definition

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:

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

若 F 有密度 f, 则其联合密度为

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

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

样本既然是随机变量, 就有一定的概率分布, 这个概率分布就叫作**样本分布**. 样本分布是样本所受随机性影响的最完整的描述.

要决定样本分布, 就要根据观察值的具体指标的性质 (这往往涉及有关的专业知识), 以及对抽样方式和对试验进行的方式的了解, 此外常常还必须加一些人为的假定. 下面看一些例子:

↑Example

一大批产品共有 N 个, 其中废品 M 个, N 已知, 而 M 未知. 现在从中抽出 n 个加以检验, 用以估计 M 或废品率 $p = M/N$.

(1) 有放回抽样, 即每次抽样后记下结果, 然后将其放回去, 再抽第二个, 直到抽完 n 个为止. 求样本分布.

(2) 不放回抽样, 即一次抽一个, 依次抽取, 直到抽完 n 个为止. 求样本分布.

↓Example

解: (1) 在有放回抽样情形, 每次抽样时, N 个产品中的每一个皆以 $1/N$ 的概率被抽出, 此时 $P(X_i = 1) = M/N$, $P(X_i = 0) =$

$(N-M)/N$, 故有

$$P(X_1=x_1,\cdots,X_n=x_n)=\left(\frac{M}{N}\right)^a\left(\frac{N-M}{N}\right)^{n-a}, \tag{4.4}$$

当 $x_1,\cdots x_n$都为 0或 1, 且 $\displaystyle\sum_{i=1}^n x_i=a$ 时为上述结果 (其余情形为 0).

(2) 若采取不放回抽样, 则计算要复杂的多, 读者可作为练习, 现将结果给出如下: 记 $\sum_{i=1}^n x_i=a$, 利用概率乘法公式易求

$$\begin{aligned}&P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)\\&=\frac{M}{N}\cdot\frac{M-1}{N-1}\cdots\frac{M-a+1}{N-a+1}\cdot\frac{N-M}{N-a}\cdots\frac{N-M-n+a+1}{N-n+1},\end{aligned} \tag{4.5}$$

当 $x_1,\cdots x_n$都为 $0,1$, 且 $\displaystyle\sum_{i=1}^n x_i=a$ 时为上述结果 (其余情形为 0) .

上述计算之所以复杂, 是因为在不放回情形, 样本 $X_1,\cdots,X_n$ 不是相互独立的, 样本分布是利用乘法公式, 通过条件概率计算出来的.

而在有放回的情形, 样本 $X_1, \cdots, X_n$ 是独立同分布的, 因此要简单得多.

当 n/N 很小时, (4.5) 和 (4.4) 差别很小. 因而当 n/N 很小时可把上例中的无放回抽样近似当作有放回抽样来处理.

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4.2.3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.

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.

↑Example

↓Example

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

$$f(x_1, \cdots, x_n) = (\sqrt{2\pi}\sigma)^{-n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2\} \tag{4.6}$$

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4.2.3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 (4.6). 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4.2.3的 $N(a, \sigma^2)$ 模型中, 有了样本 $X_1, \cdots, X_n$, 并规定分布 (4.6) 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4.2.3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4.2.3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 “$a \leq 1$” 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4.2.3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. 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:

由样本算出的量是**统计量**, 或曰,**统计量**是样本的函数.

Definition

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 σ^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**:

$$m_{\frac{1}{2}}=\begin{cases}X_{(\frac{n+1}{2})} & \text{当 n 为奇数}\\ \frac{1}{2}[X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}] & \text{当 n 为偶数}\end{cases}\tag{4.7}$$

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.

5. 经验分布函数: 假设总体分布 F 有矩, 由于不知道 F, 也就不知道矩, 现从该总体中抽出样本 $X_1, \cdots, X_n$, 我们可以使用如下函数估计 F:

$$F_n(x) = \{X_1, \cdots, X_n \text{中} \leq x \text{的个数}\}/n$$

它称为样本 $X_1, \cdots, X_n$ 的经验分布函数.

4.3.3 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 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.

证: (1) 由推论 5.5.2 立得.

(2) 设

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

为一正交阵 (这一正交阵的存在性由 Schmidt 正交化方法保证), 作正交变换 $Y = AX$, 故有

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i = \sqrt{n}\bar{X},$$

由正交变换保持向量长度不变可知

$$Y_1^2+\cdots+Y_n^2=X_1^2+\cdots+X_n^2.$$

所以

$$(n-1)S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2=\sum_{i=1}^{n}Y_i^2-Y_1^2=\sum_{i=2}^{n}Y_i^2. \tag{4.8}$$

由定理**??** 可知 $Y_i\sim N(\mu_i,\sigma^2),\ i=2,\cdots,n$. 再由 A 的行向量正交性可知

$$\mu_i=a\sum_{k=1}^{n}a_{ik}=\sqrt{n}a\cdot\sum_{k=1}^{n}\frac{1}{\sqrt{n}}\cdot a_{ik}=0. \tag{4.9}$$

以及

$$\begin{aligned} Cov(Y_i, Y_j) &= E[(Y_i - EY_i)(Y_j - EY_j)] = E\left[\sum_{k=1}^n a_{ik}(X_k - a) \cdot \sum_{l=1}^n a_{jl}(X_l - a)\right] \\ &= \sum_{k=1}^n \sum_{l=1}^n a_{ik}a_{jl}E[(X_k - a)(X_l - a)] = \sum_{k=1}^n \sum_{l=1}^n a_{ik}a_{jl}\delta_{kl}\sigma^2 \\ &= \sigma^2 \sum_{k=1}^n a_{ik}a_{jk} = \begin{cases} \sigma^2 & \text{当 } i = j, \\ 0 & \text{当 } i \neq j. \end{cases} \end{aligned}$$

此处 $\delta_{kl} = 1$, 当 $k = l$; 否则为 0. 因此 $Y_2, \cdots, Y_n$ i.i.d. $\sim N(0, \sigma^2)$. 故 $Y_i/\sigma \sim N(0,1)$, $i = 2, \cdots, n$, 因此由 (4.8) 得

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=2}^n (Y_i/\sigma)^2 \sim \chi^2_{n-1}.$$

(3) 由上述 (2) 的证明中可知 $Y_1, Y_2, \cdots, Y_n$ 相互独立, S^2 只和 $Y_2, \cdots, Y_n$ 有关, $\bar{X}$ 只和 Y_1 有关, 因此 $\bar{X}$ 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.

4.3.4 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 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 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

$$\frac{\bar{X} - \bar{Y} - (a_1 - a_2)}{\sigma}\sqrt{\frac{mn}{m+n}} \sim N(0,1). \tag{4.10}$$

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

$$\frac{(m-1)S_1^2 + (n-1)S_2^2}{\sigma^2} \sim \chi_{n+m-2}^2. \tag{4.11}$$

再由 (4.10) 和 (4.11) 中 $(\bar{X}, \bar{Y})$ 与 (S_1^2, S_2^2) 相互独立, 由定义可知

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

推论 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2所述.

证: 由注 5.4.3 可知 $(m-1)S_X^2/\sigma_1^2 \sim \chi_{m-1}^2$, $(n-1)S_Y^2/\sigma_2^2 \sim \chi_{n-1}^2$,

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.

下列这一推论给出了服从指数分布随机变量的线性函数的分布与 χ^2 分布的关系. 这在指数分布总体的区间估计和假设检验问题中有重要应用.

推论 4. 设 $X_1, X_2, \cdots, X_n$ *i.i.d.* 服从指数分布: $f(x,\lambda) = \lambda e^{-\lambda x} I_{[x>0]}$, 则有

$$2\lambda n\bar{X} = 2\lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n}.$$

证: 首先证明 $2\lambda X_1 \sim \chi^2_2$. 因为

$$F(y) = P(2\lambda X_1 < y) = P\left(X_1 < \frac{y}{2\lambda}\right) = \int_0^{\frac{y}{2\lambda}} \lambda e^{-\lambda x} dx,$$

所以

$$f(y) = F^{'}(y) = \begin{cases} \frac{1}{2}e^{-\frac{y}{2}} & \text{当 } y > 0 \\ 0 & \text{当} y \leq 0. \end{cases}$$

因此 $f(y)$ 即为自由度为 2 的 χ^2 密度, 即 $2\lambda X_1 \sim \chi^2_2$.

再利用 χ^2 分布的性质 (3), $2\lambda X_i \sim \chi^2_2$, $i = 1, 2, \cdots, n$; 又它们相互独立, 故有 $2\lambda \sum\limits_{i=1}^{n} X_i \sim \chi^2_{2n}$.

4.4 总结

获得有效数据后, 统计推断问题可以按照如下的步骤进行:

1. 确定用于统计推断的合适统计量;
2. 寻求统计量的精确分布; 在统计量的精确分布难以求出的情形, 可考虑利用中心极限定理或其它极限定理找出统计量的极限分布.
3. 基于该统计量的精确分布或极限分布, 求出统计推断问题的精确解或近似解.
4. 根据统计推断结果对问题作出解释.

其中第二步是最重要, 但也是最困难的一步. 正态总体下样本均值和样本方差的分布, 在寻求与正态变量有关的统计量精确分布时, 起着十分重要作用. 尤其在后面两章中求区间估计和假设检验问题时可以看得十分清楚.

第六章：参数估计

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第六章：参数估计

参数估计问题:

- 总体: $X \sim f_\theta(x)$,f 形式已知, $\theta = (\theta_1, \ldots, \theta_k)$ 为未知参数
- 样本: $X_1, \ldots, X_n$

利用样本对参数 θ 的作出估计或估计它们的某个已知函数 $g(\theta)$.

- **点估计**: 用样本的一个函数 $T(X_1, \ldots, X_n)$ 去估计 $g(\theta)$
- **区间估计**: 用一个区间 (区域) 去估计 $g(\theta)$

6.1 点估计

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 θ 的估计值. 为这样特定目的而构造的统计量 $\hat{\theta}$ 叫做 θ 的**估计量**. 由于参数 θ 是数轴上的一个点, 用 $\hat{\theta}$ 估计 θ, 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:

6.1.1 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, 进而得到 θ 的估计. 介绍如下: 假设总体 X 包含 k 个未知参数 $\theta = (\theta_1, \cdots, \theta_k)$, 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, 有些情况下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.

↑Example

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。

↓Example

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 $\hat{p} = \bar{X}$.

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。

↑Example

↓Example

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 $\hat{\sigma}^2 = S^2$.

6.1.2 最大似然估计方法

最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:

设样本 $X = (X_1, \ldots, X_n)$ 有概率函数

$$f(x;\theta) = f(x;\theta_1, \cdots, \theta_k)$$

这里参数 $\theta = (\theta_1, \cdots, \theta_k) \in \Theta$，$x = (x_1, \ldots, x_n)$ 为样本 X 的观察值. 当固定 x 时把 $f(x;\theta)$ 看成为 θ 的函数，称为似然函数，常记为 $L(x;\theta)$ 或 $L(\theta)$.

Definition

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**

使得能观察到 x 的可能性 $L(x;\theta)$ 最大的 θ 值, 看起来应该最像未知的 θ。这个 θ 的值即称为 θ **最大似然估计值**(看上去最有可能的)。我们先看一个例子:

↑Example

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.

↓Example

解: 先将问题一般化. 设池中有 N 条鱼, 其中 r 条做好记号. 鱼在鱼池里均匀. 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 的最可能的估计即最大似然估计为

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

其中 ⌈ ⌉ 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.

现给出最大似然估计的一般性定义:

设 $X = (X_1, \cdots, X_n)$ 为从具有概率函数 $f_\theta(x)$ 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

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

Definition

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

求最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,

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

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。

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 的最大似然估计量。

↑Example

↓Example

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

$$\begin{cases} \frac{\partial l(a,\sigma^2)}{\partial a} = 0 \\ \frac{\partial l(a,\sigma^2)}{\partial \sigma^2} = 0 \end{cases} \Rightarrow \begin{cases} a = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \\ \sigma^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - a)^2 \end{cases}$$

容易验证此驻点是唯一的最大值点，因此得到 a, σ^2 的最大似然估计量:

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

设总体 X 服从 $[a,b]$ 上的均匀分布, $a<b$, 求参数 a,b 的最大似然估计.

↑Example

↓Example

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 的最大似然估计量为 $\hat{a}=X_{(1)},\hat{b}=X_{(n)}$.

↑Example

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 的最大似然估计量.

↓Example

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 $b = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{(1)})$，容易验证此

解是最大值点。从而得到 a, b 的最大似然估计量:

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

↑Example

设总体 $X_1, \ldots, X_n$ 服从 0-1 分布 $B(1,p)$, $0<p<1$, 求参数 p 的最大似然估计.

↓Example

解: 似然函数为

$$L(p)=\prod_{i=1}^{n} P(X_i=x_i)=p^{\sum x_i}(1-p)^{n-\sum x_i}$$

从而令 $\frac{\partial log L(p)}{\partial p}=0$ 得到

$$\frac{\sum x_i}{p}=\frac{n-\sum x_i}{1-p}$$

因此 p 的似然估计为

$$\hat{p}=\frac{\sum_{i=1}^{n} X_i}{n}.$$

↑Example

设总体 $X_1,\ldots,X_n$ 服从柯西分布 $f(x)=\frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, x\in R,\theta\in R$, 求参数 θ 的最大似然估计.

↓Example

解: 因为柯西分布不存在矩, 因此矩方法不适用. 其对数似然函数为

$$l(\theta)=logL(\theta)=\sum_{i=1}^{n}logf(x_i)=\sum_{i=1}^{n}[-log\pi-log(1+(x_i-\theta)^2)$$

从而令 $\frac{\partial l(\theta)}{\partial\theta}=0$ 得到

$$\sum_{i=1}^{n}\frac{2(x_i-\theta)}{1+(x_i-\theta)^2}=0$$

此方程没有显式解, 可以使用数值方法求解. 使用起来不太方便, 因此在应用中, 考虑到柯西分布的对称性, 使用样本中位数来估计 θ.

$\overline{\uparrow}$Example

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

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

$\underline{\downarrow}$Example

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 的最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 的最大似然估计为

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

再由 $\theta = \frac{1-\sqrt{1-2\eta}}{2}$ 得到 θ 的最大似然估计为

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

6.1.3 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. 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** (Consistent Estimator)。

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，最大似然估计量在很一般的条件下也是满足相合性的。

2. 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 (Unbiased Estimator)** 。无偏性的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 的最大似然估计量 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 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.

3. 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

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

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。

设 $X_1, \cdots, X_n$ 为从如下分布中抽取的简单样本，试证明样本方差为总体方差的无偏估计.

↑Example

↓Example

证: 显然

$$
\begin{aligned}
ES^2 &= E\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(E(X_i-EX_i+EX_i-\bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}[(E(X_i-EX_i)^2-E(EX_i-\bar{X})^2] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}[\sigma^2-\sigma^2/n)^2=\sigma^2.
\end{aligned}
$$

设总体 X 服从 $(0,\theta)$ 上的均匀分布, $0<\theta$, 求参数 θ 的最大似然估计是否为无偏估计.

↑Example

↓Example

解: 易得似然函数为

$$L(\theta)=\frac{1}{\theta^n}\prod_{j=1}^{n}I(0\le x_j\le\theta)=\frac{1}{\theta^n}I(0\le x_{(1)}\le x_{(n)}\le\theta).$$

于是似然函数 $L(\theta)$ 在 $\theta=x_{(n)}$ 时取到最大值. 而 $X_{(n)}$ 的密度函数为

$$f(t)=\frac{nt^{n-1}}{\theta^n}I(0<t<\theta).$$

因此

$$EX_n=\int_0^\theta tf(t)dt=\frac{n}{n+1}\theta.$$

即 θ 的最大似然估计量 $X_{(n)}$ 不是 θ 的无偏估计, 但 $\frac{n+1}{n}X_{(n)}$ 为 θ 的无偏估计量.

设 $X_1,\ldots,X_n$ 来自均值为 μ, 方差为 σ^2 的总体分布的简单样本, $\omega_1,\ldots,\omega_n$ 为已知的非负权值, 且满足 $\sum\omega_i=1$, 试比较 μ 的两个估计估计 $\bar{X}$ 和 $\sum_{i=1}^n\omega_iX_i$.

↑Example

↓Example

解: 因为

$$Var(\bar{X})=\frac{\sigma^2}{n},\qquad Var(\sum\omega_iX_i)=\sum_{i=1}^n\omega_i^2\sigma^2,$$

所以

$$Var(\sum\omega_iX_i)\geq Var(\bar{X})$$

且等号成立当且仅当 $\omega_i=\frac{1}{n}$.

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”。

↑Example

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 $\hat{\theta}_M$ 和最大似然估计量 $\hat{\theta}_L$，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.

↓Example

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。

第六讲：参数估计

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第六讲：参数估计

6.2 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的可信程度. 这种形式的估计称为**区间估计**.

↑Example

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.

↓Example

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.

6.2.1 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 (一维未知)θ 的区间估计, 就是要

- **寻求统计量 $\underline{\theta}(X_1, \cdots, X_n) < \bar{\theta}(X_1, \cdots, X_n)$ 所构成的区间 $[\underline{\theta}, \bar{\theta}]$.**
- **该区间满足一定的要求**

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:

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 θ 的置信水平为 $1-\alpha$ 的置信区间。

Definition

置信区间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 $a \le S(T, g(\theta)) \le b$ 要能表示成等价的形式 $A \le g(\theta) \le B$，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

$$P(\omega_{1-\alpha/2} \le S(T, g(\theta)) \le \omega_{\alpha/2}) = 1-\alpha$$

由 3 我们就可以得到所求的置信区间.

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。

↑Example
↓Example

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

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

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

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。

↑Example

↓Example

解： 方法完全类似于前面的例子，由于 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ 的估计分别为 $\bar{X}, \bar{Y}, S_X^2, S_Y^2$，且注意到 $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m})$，$(n-1)S_X^2/\sigma_1^2 \sim \chi_{n-1}^2$ 以及 $(m-1)S_X^2/\sigma_1^2 \sim \chi_{m-1}^2$，结合两组样本的独立性可知

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

从而可得 σ_X^2/σ_Y^2 的置信区间. 对 $\mu_1 - \mu_2$ 的置信区间, 当 σ_X^2, σ_Y^2 已知或者相等但未知情形, 容易得到其置信区间; 当两者不全已知且不相等时, 不存在 $\mu_X - \mu_Y$ 的精确置信区间 (Behrens-Fisher problem).

2. 大样本法

大样本法就是利用极限分布，以建立枢轴变量。通过以下例子说明:

↑Example

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。

↓Example

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:

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

6.2.2 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.

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**.

Definition

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。

6.2.3 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。

↑Example

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。

↓Example

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 μ 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 $n \geq \left(\frac{2\times1.96\times0.1}{0.05}\right)^2 = 61.4656$. 即为达到要求至少需要测量 62 次。

第七讲：假设检验

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第七讲：假设检验

7.1 基本概念和问题的提法

- (统计) 假设: 在数理统计中, 关于总体分布的概率性质的假定. 例如假设正态总体, 二项总体等, 或者二项总体中成功概率 $p \leq 0.5$ 等等.
- (统计) 检验: 使用样本对所作出的假设进行检查的方法和过程.

7.1.1 基本概念

假设检验问题就是研究如何根据抽样后获得的样本来检查抽样前所作假设是否合理.

首先, 由一个例子引出一些基本概念.

↑Example

某厂产品出厂检验规定: 每批产品次品率 p 不超过 4% 才能出厂。现从某批产品 10000 件中任意抽查 12 件发现 4 件次品, 问该批产品能否出厂? 若抽得结果是 1 件次品呢?

↓Example

解: 若以 p 表示此批产品的次品率，则问该批产品能否出厂等价于即要检验次品率 p 是否不超过 4%。我们假设“$p \le 4\%$”，并记 Y 为 12 件中的次品数，由于总产品数很大，故可以认为 $Y \sim B(12, p)$，此时当 $p \le 0.04$ 时，

$$P(Y = 4) = \binom{12}{4} p^4 q^8 < \binom{12}{4} 0.04^4 0.96^8 = 0.000914$$

这是一个小概率事件，即当 $p \le 0.04$ 时，12 件产品中有 4 件是次品的概率不到 1/1000，这样的事件在一次试验中几乎是不可能发生的，但确实发生了 (我们观察到了 4 件次品), 因此更倾向于怀疑假设“$p \le 0.04$”的正确性，即认为它不成立。而由于

$$P(Y = 1) \le \binom{12}{1} 0.04^1 0.96^{12} = 0.306$$

即此时当假设“$p \le 0.04$”成立时，“12 个产品中有一个次品”这一事件的概率最大为 0.306，这个事件不是小概率事件。因此我们没有足够的证据支持原假设不成立这一说法。

↑Example

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.

↓Example

在这个问题中,

统计假设: 罐装饮料容量 $X \sim N(\mu, 10^2)$.

问题: 根据样本来在 "$\mu = 500$" 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**不能拒绝** 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**.

决策 \ 事实	H_0 成立	H_1 成立
不拒绝 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)**
- **"实际上 H_0 不成立但是它没有被拒绝" 这样一类错误**为 **第 II 类错误 (存伪)**.

而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). 因此，**这种在只限制第一类错误的原则下的检验方法，就称为“显著性检验”**(Significance Test)。

具体地, 给定一个允许的犯第一类错误概率的最大值 α, 选取 τ 使得

$$P_{H_0}(T < \tau) \leq \alpha$$

连续场合下, 等号可以达到. 这样的 τ 可以通过 T 在 H_0 下的分布及上式条件下求得.

称 α 为**显著性水平**. 通常将 α 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:

1. 根据问题, 提出假设检验问题

$$H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1.$$

   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**.

2. 根据参数的估计方法构造一个适当的**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.

3. 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 τ 为 **临界值**.

4. 对任意的 $\theta \in \Theta_0$, 犯第 I 类错误的概率 $P_\theta(T(X_1, \cdots, X_n) \in A)$ 小于或等于某个指定正的常数 α), 则称 α 为**显著性水平**.

5. 结合 T 在 H_0 下的分布, 定出 A.

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.

称 $\beta(\theta) = P_\theta$（H_0 被拒绝）为检验的**功效函数**. Definition

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.

显著性检验方法由于只控制第一类错误, 因此保护了原假设不被轻易拒绝, 从而原假设和对立假设地位不相同!

7.1.2 原假设的提法

在有时候需要自己判断如何提假设检验问题. 在建立原假设时有两个原则。

原则一：将受保护的对象置为零假设. 将已经存在的事实, 或者错误拒绝会带来很大后果的事情作为原假设. 例如, 司法上的无罪推断. 这样做大大地有利于保护公民的利益. 又若对新药的批准, 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. 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.

原则三：假设检验的“拒绝零假设”结果比“不能拒绝零假设”更有保证. 当我们没有任何知识来选择零假设时候，提一个零假设使得使用某个检验方法得到的结果为“拒绝零假设”. 因为零假设不被轻易拒绝，其犯错的概率不超过 α，因此得到“拒绝零假设”的结论比得到”不能拒绝零假设”的结论更有保证.

7.1.3 检验统计量的选取及假设检验的步骤

通过解答前例来说明假设检验的步骤.

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?

↑Example

↓Example

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 $H_0: \ \theta \leq \theta_0 \leftrightarrow H_1: \ \theta > \theta_0$

(4) $H_0: \ \theta = \theta_0 \leftrightarrow H_1: \ \theta < \theta_0$ 或者 $H_0: \ \theta \geq \theta_0 \leftrightarrow H_1: \ \theta < \theta_0$

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.

对上述这些假设, 我们总结一下显著性检验的一般步骤, 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.

第七讲：假设检验

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第七讲：假设检验

7.2 一样本和两样本总体参数检验

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).

7.2.1 一样本正态总体参数检验

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. 则可能考虑的参数有均值 μ 和方差 σ^2:

(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**

$$Z = Z(X_1, \cdots, X_n) = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}$$

注意到当 H_0 成立时, $U \sim N(0, 1)$, $|Z|$ 应该较小, 反之当 $|U|$ 的观测值 $z(x_1, \cdots, x_n)$ 较大时, 不利于零假设 H_0 应该拒绝之. 所以选拒绝域形如

$$\{|Z| > \tau\}.$$

要求显著性水平为 α, 即

$$P_{H_0}(|Z| > \tau) = \alpha,$$

解得 $\tau = z_{\alpha/2}$. 于是检验的拒绝域为

$$\{|Z| > u_{\alpha/2}\}.$$

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**右侧假设**

$$H_0:\ \mu=\mu_0 \leftrightarrow H_1:\ \mu>\mu_0 \quad \text{或者} \quad H_0:\ \mu\le\mu_0 \leftrightarrow H_1:\ \mu>\mu_0$$

仍然用统计量 Z, 由于 Z 大时不利于 H_0, 取拒绝域为

$$\{Z>u_\alpha\}.$$

而检验另一个**左侧假设**

$$H_0:\ \mu=\mu_0 \leftrightarrow H_1:\ \mu<\mu_0 \quad \text{或者} \quad H_0:\ \mu\le\mu_0 \leftrightarrow H_1:\ \mu<\mu_0$$

的拒绝域为

$$\{Z<-u_\alpha\}.$$

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 Z **检验.**

↑Example

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?

↓Example

解: 这是方差已知时关于均值 μ 的假设检验问题,

$$H_0 : \mu = 3 \leftrightarrow H_1 : \mu \neq 3.$$

取检验统计量为 $Z=\sqrt{n}(\bar{X}-3)/0.1$, 检验的拒绝域为 $|Z|>u_{\alpha/2}$. 由样本算得检验统计量的值为 $z\approx 2.16$, 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.**

↑Example

对正态总体 $N(\mu,\sigma^2)$(其中 σ^2 已知) 下的假设检验问题 $H_0:\mu\geq\mu_0\leftrightarrow H_1:\mu<\mu_0$，如果我们还要求“犯第二类错误的概率要小于指定的 $\beta>0$”该怎么办?

↓Example

解：根据功效函数和两类错误的定义，知道等价的要求

$$\beta_\phi(\mu)\geq 1-\beta,\quad \mu<\mu_0 \tag{7.1}$$

但是，当 $\mu<\mu_0$ 但 μ 接近 μ_0 时，$\beta_\phi(\mu)\approx\alpha$，而因为 α,β 一般都很小，因此一般有 $\alpha<1-\beta$，这就看出要求 (7.1) 无法达到。我们

只能放松一些，要求对某个指定的 $\mu_1 < \mu_0$，有

$$\beta_\phi(\mu) \geq 1 - \beta, \quad \mu < \mu_1 \tag{7.2}$$

因为 $\beta_\phi(\mu)$ 为 μ 的减函数，因此等价于要求

$$\beta_\phi(\mu_1) \geq 1 - \beta$$

此即

$$\Phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} - u_\alpha\right) \geq 1 - \beta$$

等价的得到

$$n \geq \sigma^2 (u_\alpha + u_\beta)^2 / (\mu_0 - \mu)^2$$

也即要满足题目中的要求，样本大小至少要达到上式右边那么大。□

(2) 方差未知时均值的检验

考虑检验

$$H_0 : \mu = \mu_0 \leftrightarrow \mu \neq \mu_0,$$

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, 见表 7.2.1 中.

(例7.2.1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?

↑Example

↓Example

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 χ^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, 参看下表 7.2.1.
上述检验称为 χ^2 检验.

表 7.2.1: 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)	$Z=\sqrt{n}(\bar{X}-\mu_0)/\sigma$	$N(0,1)$	$\begin{cases} \lvert Z\rvert > u_{\alpha/2} \\ Z > u_\alpha \\ Z < -u_\alpha \end{cases}$
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^2_n	$\begin{cases} \chi^2 > \chi^2_n(\alpha/2)\text{或者}\chi^2 < \chi^2_n(1-\alpha/2) \\ \chi^2 > \chi^2_n(\alpha) \\ \chi^2 < \chi^2_n(1-\alpha) \end{cases}$
σ^2 (μ 未知)	$\chi^2=\frac{1}{\sigma_0^2}\sum_{i=1}^n(X_i-\bar{X})^2$	χ^2_{n-1}	$\begin{cases} \chi^2 > \chi^2_{n-1}(\alpha/2)\text{或者}\chi^2 < \chi^2_{n-1}(1-\alpha/2) \\ \chi^2 > \chi^2_{n-1}(\alpha) \\ \chi^2 < \chi^2_{n-1}(1-\alpha) \end{cases}$

†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$\sigma^2<\sigma_0^2$.

(例7.2.1续) 在水平 0.1 下能否认为铁钉的标准差大于 0.1 厘米?

↑Example
↓Example

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.

7.2.2 两样本正态总体的情形

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.

↑Example

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
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?

↓Example

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.

↑Example

在例7.2.2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

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \leftrightarrow H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1.$$

↓Example

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 7.2.2: 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)	$Z=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$	$N(0,1)$	$\begin{cases} \lvert Z\rvert>u(\alpha/2) \\ Z>u(\alpha) \\ Z<-u(\alpha) \end{cases}$
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

7.2.3 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**

$$\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$$

- **数据对之间通常可以认为是独立的**
- **数据对内两个观测通常不独立**

比如一个病人在用药前后测得的指标分别为 X 和 Y

当数据是**连续数据**时候，通常对数据对内取差，构造一个新的总体 $Z = Y - X$ 及样本 $Z_1 = X_1 - Y_1, \cdots, Z_n = X_n - Y_n$，通常假设 Z 服从正态，则相应的假设检验转为一样本正态检验问题!

7.2.4 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

$$\{|T| > u_{\alpha/2}, \quad \{T > u_\alpha\} \quad \text{和} \quad \{T < -u_\alpha\}$$

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?

↑Example
↓Example

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.

7.2.5 置信区间和假设检验之间的关系

置信区间和假设检验之间有着明显的联系。我们首先考虑置信区间和双边检验之间的关系。设 $X_1, \cdots, X_n$ 为从总体 $F(x;\theta)$ 中抽取的样本，参数 θ 的 $1-\alpha$ 置信区间为 $[\underline{\theta}, \bar{\theta}]$, 即

$$P(\underline{\theta} \leq \theta \leq \bar{\theta}) \geqq 1-\alpha$$

而对假设 $H_0: \theta = \theta_0 \leftrightarrow H_1: \theta \neq \theta_0$，在原假设之下，有

$$P(\underline{\theta} \leq \theta_0 \leq \bar{\theta}) \geqq 1-\alpha$$

等价于

$$P(\theta_0 > \bar{\theta}) + P(\theta_0 < \underline{\theta}) \leq \alpha$$

按显著性检验的定义，即得其检验为

$$\phi: \text{当 } \underline{\theta} \leq \theta_0 \leq \bar{\theta} \text{ 时，接受} H_0, \text{不然就拒绝}$$

反过来讲, 如果假设 $H_0: \theta = \theta_0 \leftrightarrow H_1: \theta \neq \theta_0$ 检验的接受域有形式

$$\underline{\theta}(x_1, \cdots, x_n) \leq \theta_0 \leq \bar{\theta}(x_1, \cdots, x_n)$$

即有

$$P(\underline{\theta} \leq \theta_0 \leq \bar{\theta}) \geqq 1 - \alpha$$

由 θ_0 的任意性，知对任意的 θ，有

$$P(\underline{\theta} \leq \theta \leq \bar{\theta}) \geqq 1 - \alpha$$

即：为求出参数 θ 的 $1-\alpha$ 置信区间，我们可以先找出 θ 的双边检验 $H_0: \theta = \theta_0 \leftrightarrow H_1: \theta \neq \theta_0$ 的检验函数，则其接受域就是参数 θ 的 $1-\alpha$ 置信区间。反过来，为求假设 $H_0: \theta = \theta_0 \leftrightarrow H_1: \theta \neq \theta_0$

的检验，我们可以先求出参数 θ 的 $1-\alpha$ 置信区间，则就是该假设的接受域。

类似地，置信水平为 $1-\alpha$ 的单侧置信区间 $(\underline{\theta},\infty)$ (或者 $(-\infty,\bar{\theta})$) 与显著性水平为 α 的右 (或者左) 边检验问题 $H_0:\theta\leq\theta_0 \leftrightarrow H_1:\theta>\theta_0$ (或者 $H_0:\theta\geq\theta_0 \leftrightarrow H_1:\theta<\theta_0$)，也有类似的对应关系。

第七讲：假设检验

陈昱

cyu@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~cyu/

论坛 http://fisher.stat.ustc.edu.cn

第七讲: 假设检验

7.3 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

$$H_0 : X\text{服从某种分布}F$$

可以采用 Karl Pearson 提出的 χ^2 拟合优度检验.

基本想法: 基于样本得到 F 的估计 $\hat{F}_n$, 计算某种偏差 $D(\hat{F}_n, F)$, 例如 $sup_{x\in R}|\hat{F}_n(x) - F(x)|$. 当 H_0 正确时, 由于 $\hat{F}_n$ 是 F 的相合估计, 偏差 $D(\hat{F}_n, F)$ 应该很小.

Karl Pearson 对离散分布 F 提出一种检验方法, 即拟合优度检验方法或者称为 Pearson 卡方检验方法.

7.3.1 离散总体情形

(1) 理论总体分布不含未知参数的情形

设某总体 X 服从一个离散分布,

X	a_1	$\dots$	a_k
P	p_1	$\dots$	p_k

$p_1, \cdots, p_k$ 完全已知. 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:

$$H_0 : P(X = a_1) = p_1, \cdots, P(X = a_k) = p_k.$$

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. Pearson 提出检验统计量

$$T = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \sum \frac{(O-E)^2}{E}.$$

可以严格地证明, 在一定的条件下, 当 H_0 成立时, T 的极限分布就是自由度为 $k-1$ 的 χ^2 分布.

拒绝域: $T > \chi^2_\alpha(k-1)$

下面给出一个例子来说明拟合优度检验的应用.

↑Example

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?

↓Example

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 $\chi^2_5(0.2) \approx 7.29$ 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.

孟德尔 (Mendel) 豌豆杂交试验。纯黄和纯绿品种杂交，因为黄色对绿色是显性的，在 Mendel 第一定律 (自由分离定律) 的假设下，二代豌豆中应该有 75%是黄色的，25%是绿色的。在产生的 $n = 8023$ 个二代豌豆中，有 $n_1 = 6022$ 个黄色，$n_2 = 2001$ 个绿色。我们的问题是检验这些这批数据是否支持 Mendel 第一定律，要检验的假设是

$$H_0: \quad \pi_1 = 0.75, \quad \pi_2 = 0.25$$

↑Example

↓Example

解: 在 Mendel 第一定律 (H_0) 下，黄色和绿色的个数期望值为

$$\mu_1 = n\pi_1 = 8023*0.75 = 6017.25, \ \mu_2 = n\pi_2 = 8023*0.25 = 2005.75$$

则 Pearson χ^2 统计量为

$$Z=\sum \frac{(O-E)^2}{E}=(6022-6017.25)^2/6017.25+(2001-2005.75)^2/2005.75=0.015$$

自由度 $df = 1$，$p - value$ 为 0.99996. 因此可以认为这些数据服从 Mendel 第一定律。Fisher 基于 Mendel 的这些数据，发现其数据与

理论值符合的太好，$p - value = 0.99996$，但这么好的拟合在几千次试验中才发生一次，因而 Fisher 断定数据可能有伪造的嫌疑。

(2) 理论总体分布含若干未知参数的情形 设某总体 X 服从一个离散分布，

X	a_1	$\dots$	a_k
P	p_1	$\dots$	p_k

$p_i = p_i(\theta_1, \dots, \theta_r), i = 1, \dots, k$ 依赖于 r 个未知参数 $\theta_1, \dots, \theta_r$. 此时理论频数 np_i 一般也与这些参数有关, 从而使用最大似然估计代替这些参数以得到 p_i 的最大似然估计 $\hat{p}_i$, 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 $k-1-r$.

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?

↑Example

↓Example

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.

7.3.2 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

$$p_{ij} = P(\text{属性 A, B 分别处于水平}i,j), \tag{7.1}$$

$$u_i = P(\text{属性 A 有水平}i), \tag{7.2}$$

$$v_i = P(\text{属性 B 有水平}j) \tag{7.3}$$

则零假设等价于

$$H_0: p_{ij} = u_i v_j \quad \forall i,j$$

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.

↑Example

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)

↓Example

解: 这是一个齐一性检验问题. 检验统计量 χ^2 的观测值为 0.2524, 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.

7.3.3 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

$$\begin{aligned} p_j &= P_{H_0}(a_{j-1} < X \le a_j) \\ &= F_0(a_j;\theta_1,\cdots,\theta_r) - F_0(a_{j-1};\theta_1,\cdots,\theta_r), j=1,\cdots,k. \end{aligned}$$

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.

↑Example

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	16	14	12	17

可否在显著性水平 0.05 下认为该总体服从正态分布?

↓Example

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 $\sigma^2 = \frac{99}{100} \times 1.282^2 = 1.622$ 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 9.25, 与自由度为 6-1-2=3 的 χ^2 分布的上 0.05 分位数 $\chi^2_3(0.05) \approx 7.81$ 比较可以拒绝零假设, 即可以在显著性水平 0.05 下认为该总体不服从正态分布.

P 值　若检验的拒绝域为 $T(\mathbf{X}) > \tau$, 对两组不同的样本 $\mathbf{X}_1$ 和 $\mathbf{X}_2$, 若它们均落在拒绝域:

$$T(\mathbf{X}_1) > \tau, \qquad T(\mathbf{X}_2) > \tau$$

则它们否定原假设的程度一样吗? 如何区分这个差异?

$$P\text{值} = P(\text{在}H_0\text{ ,得到如检验统计量}T(\mathbf{X})\text{的值}T(\mathbf{x})\text{这么大或者更极端})$$

从而可以通过 P 值来比较样本的支持程度.

对不同的水平 α 检验方法, 可以通过比较它们的功效 (二型错误) 来评比优劣.