教师简介

姓　　名　何志坚
职　　称　数学学院副教授
研究兴趣　统计模拟与计算、随机算法、金融工程
联系方式　hezhijian@scut.eud.cn (by email only!)
办公地址　五山校区四号楼4301
个人主页　www.hezhijian.com

教材

主要教材:

- 何春雄，龙卫江，朱锋峰。概率论与数理统计。高等教育出版社。2012.

辅助教材:

- 茆诗松，程依明，濮晓龙。概率论与数理统计教程(第二版)。高等教育出版社。2011.
- Jay L. Devore. Probability and Statistics (第5版)(影印版), 高等教育出版社, 2004.
- 薛毅，陈立萍。R语言实用教程。清华大学出版社。2014.

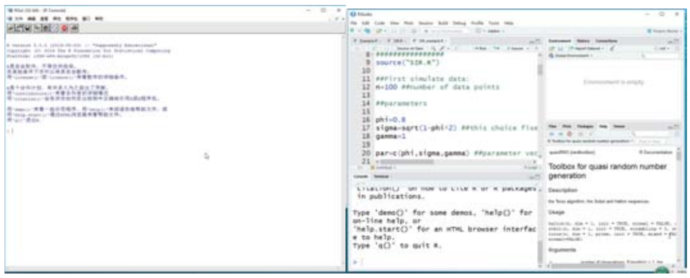
教学安排

上课时间：1-16周每周五上午，9-16周每周三下午（48学时）
上课地点：大学城校区A1-308
公共邮箱：myslideshare@163.com（密码：sharemyslide）

第一章	随机事件与概率	1-5周（10学时）
第二章	一维随机变量及其分布	6-8周（6学时）
第三章	随机向量及其分布	9-10周（6学时）
第四章	随机变量的数字特征	10-11周（6学时）
第五章	大数定律和中心极限定理	12周（4学时）
第六章	数理统计的基本概念	13周（4学时）
第七章	参数估计	14-15周（6学时）
第八章	假设检验	15-16周（4学时）
总复习	总复习	16周（2学时）

R软件（自学）

R软件下载地址：`https://www.r-project.org/`
RStudio编辑器：`https://www.rstudio.com/`



考核方式

最终成绩 = 平时成绩(30%) + 期末成绩(70%)

- 平时成绩：作业+考勤
- 期末考试：闭卷、百分制、统一出卷和改卷
- 学霸模式：如果期末考满分，则最终成绩为满分！

几点说明

- 遵守课堂纪律，请参考《学生手册》。
- 考试结束后不接受分数查询，也不接受任何求情。
- 质疑分数者可按有关规定查阅试卷。
- 欢迎找我讨论问题（课间、办公室或者邮件），但希望你的问题是"经过思考的"。在提问前，我鼓励大家事先通过网上资源或者与同学交流来寻找答案。这样可提高自学能力。
- 欢迎对本课程提宝贵意见，发送至hezhijian@scut.edu.cn

学这门课程有什么用?

- 可以凑学分，毕业?
- 可以学到有用的知识?
- 用严格的数学方法研究随机现象！

案例

- 用1部新手机充电五分钟测试，观测通话时间为1.5h
- 用50部新手机充电五分钟测试，观测50次通话时间为

$$1.7 \quad 1.8 \quad 1.3 \quad 2.1 \quad 2.3 \quad \cdots \quad 2.5$$

- 每次测试通话时间是随机的
- 如何刻画这样的规律？概率论：概率分布
- 如何基于这些实验数据做出判断？统计学：假设检验

第一章：随机事件与概率

目录

随机试验、随机现象和随机事件

- 随机试验特点：可重复性、不可预知性
- 随机试验观测到的现象为随机现象
  - 概率论与数理统计研究的对象
  - 概率论研究随机现象的统计规律
  - 数理统计研究随机现象的数据收集与分析
- 随机试验的某些可能结果组成的集合称为随机事件，简称事件

序号	随机试验	事件A	事件B
(1)	观测一部手机的通话时间	通话时间等于2h	通话时间大于2h
(2)	观测一颗骰(tóu)子的点数	点数为奇数	点数为偶数
(3)	某新型药的治疗效果	治疗有效	治疗无效

样本空间与随机事件

样本空间：随机试验所有可能结果的集合称为样本空间。常用Ω表示。
样本点：样本空间的元素称为样本点，常用ω表示。

序号	随机试验	样本空间
(1)	观测一部手机的通话时间	$\Omega = [0, \infty)$
(2)	观测一颗骰(tóu)子的点数	$\Omega = \{1, 2, 3, 4, 5, 6\}$
(3)	某新型药的治疗效果	$\Omega = \{$治疗有效, 治疗无效$\}$

有限样本空间：试验(2)和试验(3)
无限样本空间：试验(1)
随机事件是某些样本点组成的集合。

随机事件

在一次试验中，当试验结果属于事件A时，称这次试验中事件A发生。否则称事件A不发生。

必然事件	$A=\Omega$
不可能事件	$A=\emptyset$
基本事件/简单事件	$A=\{\omega\},\ \omega\in\Omega$
复合事件	$A\subseteq\Omega$

注意：基本事件是相对的，不是绝对的！

例

在一批含有20件正品，5件次品的产品中随机地抽取2件，可能结果如下：

A = {2件全是正品}

B = {只有1件是正品}

C = {2件全是次品}

- 在不计次序的假定下，A、B、C是基本事件。
- 如果考虑次序，B不再是基本事件，它可分解为B_1和B_2两个基本事件。
  B_1 = {第1次抽到正品，第2次是次品}
  B_2 = {第1次抽到次品，第2次是正品}

事件的关系

- 事件的包含：如果事件A发生，事件B一定发生。则称事件B包含事件A。记为$A\subset B$。显然$A\subset\Omega$。
- 事件的相等：$A=B\Leftrightarrow A\subset B$且$B\subset A$
- 事件的互斥：$A\cap B=\varnothing$
- 事件的对立：$A\cap B=\varnothing$且$A\cup B=\Omega$，记$A=\bar{B}$或者$B=\bar{A}$

考虑投骰子的例子：$A=$ {出现偶数点}，$B=$ {出现奇数点}，$C=$ {出现点数1和3}，$D=$ {出现点数5和6}，$E=$ {出现点数大于4}

- 事件A和B的关系？
- 事件B和C的关系？
- 事件C和D的关系？
- 事件D和E的关系？

事件的运算：交、并、差

- 事件的交(或积)：$A\cap B$或者AB，事件A和B同时发生
- 事件的并(或和)：$A\cup B$，事件A和B至少一个发生。若A和B互斥，则$A\cup B$可表示为$A+B$。
- 事件的差：$A\backslash B$，或表示为$A-B$，事件A发生但事件B不发生
- 多个事件的交$\bigcap_{i=1}^{n} A_i$，多个事件的并$\bigcup_{i=1}^{n} A_i$

▶韦恩图

例

设A、B、C为任意三个事件，写出下列事件的表达式：

- AC都发生B不发生。$AC\backslash B=A\bar{B}C$
- 恰有二个事件发生。$AB\bar{C}\cup A\bar{B}C\cup\bar{A}BC$
- 至少有一个事件发生。$A\cup B\cup C$
- 三个事件同时发生。ABC
- 三个事件都不发生。$\bar{A}\bar{B}\bar{C}$
- 三个事件不都发生。$\overline{ABC}$

事件的运算法则

对于任意三个事件A,B,C，满足下列运算：

- 交换律：$AB=BA,\ A\cup B=B\cup A$
- 结合律：$(AB)C=A(BC),\ (A\cup B)\cup C=A\cup(B\cup C)$
- 分配律：$A(B\cup C)=AB\cup AC,\ A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$
- 对偶律：$\overline{A\cup B}=\bar{A}\cap\bar{B},\ \overline{A\cap B}=\bar{A}\cup\bar{B}$

推广到多个事件：

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

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

目录

概率的统计定义

频率的定义：设事件A在n次试验中出现了r次，则比值r/n称为事件A在n次试验中出现的频率。

概率的统计定义：在同一组条件下所作的大量重复试验中，事件A出现的频率总是在区间$[0,1]$上的一个确定的常数p附近摆动，并且稳定于p（频率的稳定值），则p称为事件A的概率，记作$P(A)$。

注意：P(这里面只能为事件！)。如，$P(\text{骰子的点数}) = 1/6$是错误的。

古典概率

古典概型的随机试验要求满足下两条件:

- 有限性。只有有限多个不同的基本事件。
- 等可能性。每个基本事件出现的可能性相等。

在古典概型中，如果基本事件（样本点）的总数为n，事件A所包含的基本事件（样本点）个数为$r(r \le n)$，则定义事件A的概率$P(A)$为r/n。即

$$P(A) = \frac{r}{n} = \frac{\text{A中包含的基本事件的个数}}{\text{基本事件的总数}}$$

简单例子：抛均匀的硬币、骰子

思考：古典概型的取值范围?

例1：抽球类型

袋中有a个黄球，b个白球，从中接连任意取出k个球$(k \le a+b)$，且每次取出的球**不再放回去**，求第k次取出的球是黄球的概率?

- 事件A={第k次取出的球是黄球}
- 基本事件?
- 是否是古典概型? 有限性、等可能性?

解法一：考虑第1次到第k次的取球结果。

$$P(A) = \frac{C_a^1 A_{a+b-1}^{k-1}}{A_{a+b}^k} = \frac{a}{a+b}$$

解法二：只考虑第k次的取球结果。

$$P(A) = \frac{C_a^1}{C_{a+b}^1} = \frac{a}{a+b}$$

结论：抽签与顺序无关（如同有放回的情况）！基本事件是相对的！

例2：m个质点在n个格子中的分布问题

设有m个不同质点，每个质点都以概率$1/n$落入n个格子$(n \ge m)$的每一个之中，求下列事件的概率:

1. A: 指定m个格子中各有一个质点;
2. B: 任意m个格子中各有一个质点;
3. C: 指定的一个格子中恰有$k(k \le m)$个质点。

解:

$$P(A) = \frac{m!}{n^m}$$
$$P(B) = \frac{C_n^m m!}{n^m} = \frac{A_n^m}{n^m}$$
$$P(C) = \frac{C_m^k (n-1)^{m-k}}{n^m}$$

例3：同一天生日问题

某班级有n个人，问至少有两个人的生日在同一天的概率为多大?（假设一年有365天）

解:

记$A = \{n\text{个人中至少有两个人的生日相同}\}$。$\bar{A} = \{n\text{个人中的生日全不相同}\}$，易知

$$P(\bar{A}) = \frac{A_{365}^n}{365^n} = \frac{365!}{365^n(365-n)!}.$$

因此，

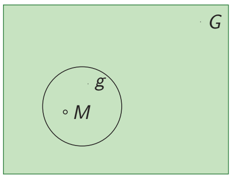
$$P(A) = 1 - P(\bar{A}) = 1 - \frac{365!}{365^n(365-n)!}.$$

注意：上述只是对$n \le 365$成立。如果$n > 365$，则显然$P(A) = 1$.

概率的几何定义

平面上有可测的区域G和g，向G中随机投掷一点M，设M必落在G内。如M落在g内的概率只与g的面积成正比，而与g的位置和形状无关。这样的随机实验，称为几何概型。点M落入G内的部分区域g的概率为：

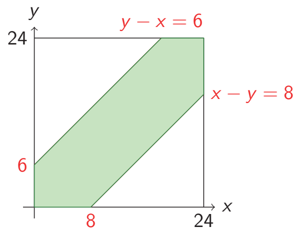
$$P = \frac{g\text{的面积}}{G\text{的面积}}$$



注意：随机投点是指M落入G内任一处均是等可能的。
思考： 几何概型的取值范围?

例1：会面问题

已知甲乙两船将在同一天的0点到24点之间随机地到达码头，该码头只有一个泊位。若甲先到达，需停靠6小时后才离开码头。若乙先到达，则要停靠8小时后才离开码头。问这两船中有船需等候泊位空出的概率?
解：设甲船到达码头的时刻是x，乙船到达码头的时刻是y，显然$0 \le x, y \le 24$。若这两船中有船需等候泊位空出，则$y - x < 6$且$x - y < 8$.



$$P(A) = \frac{24^2 - (16^2 + 18^2)/2}{24^2}$$
$$\approx 0.4965$$

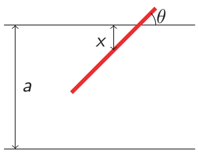
例2：蒲丰投针实验

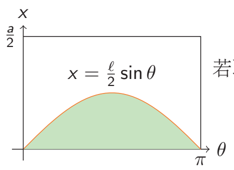
蒲丰是几何概率的开创者，并以蒲丰投针问题闻名于世，发表在其1777年的论著《或然性算术试验》中。
由于通过他的投针试验法可以利用很多次随机投针试验算出π的近似值，所以特别引人瞩目，这也是最早的几何概率问题。并且蒲丰本人对这个实验给予证明。
1850年，瑞士数学家沃尔夫在苏黎世，用一根长$36mm$的针，平行线间距为$45mm$，投掷5000次，得$\pi \approx 3.1596$。1864年，英国人福克投掷了1100次，求得$\pi \approx 3.1419$。1901年，意大利人拉泽里尼投掷了3408次，得到了准确到6位小数的π值。

图：George-Louis Leclerc de Buffon (1707.9.7-1788.4.16)，法国数学家、自然科学家。

例2：蒲丰投针实验

平面上画着一些平行线，它们之间的距离等于a,向此平面任投长度为$\ell(\ell < a)$的针，试求此针与任一平行线相交的概率。
解：设x表示针的中点到最近的一条平行线的距离，θ表示针与平行线的交角。显然$0 \le x \le a/2,\ 0 \le \theta \le \pi$. 为使针与平行线相交，必须$x \le \frac{\ell}{2}\sin\theta$.
所求的概率为：





$$P = \frac{\frac{\ell}{2}\int_0^\pi \sin\theta\, d\theta}{\frac{1}{2}\pi a} = \frac{2\ell}{\pi a}.$$

若取$\ell = a/2$，则

$$\pi = 1/P.$$

▶模拟动画

主观概率

统计界的贝叶斯学派认为：一个事件的概率是人们根据经验对该事件发生的可能性所给出的个人信念。

- 天气预报说“明天下雨的概率为90%”
- 医生说“手术成功可能性为90%”
- 我说“考试及格的可能性为99%”

注意区别“主观概率”与“主观臆造”。

概率的公理化定义

前面学了三种概率定义，各有其局限性。

- 统计概率：要求作大量重复试验。很难观测到稳定值
- 古典概率：试验结果要求有限、互不相容、等可能。不均匀的硬币
- 几何概率：落入区域G内任一点是等可能的。高手投飞镖

如何刻画更一般的情况? 1900年数学家希尔伯特(Hilbert, 1862–1943)提出要建立概率的公理化定义来解决这个问题，即以最少的几条本质特征出发去刻画概率的概念。1933年苏联数学家柯尔莫戈夫(Kolmogorov, 1903–1987)首次提出了概率的公理化定义，这个定义既概括了历史上几种概率的共同特征，又避免了各自的局限性和含混之处，不管什么随机现象，只要满足该定义的三条公理，才能说它是概率。

概率的公理化定义

样本空间Ω，事件域$\mathcal{F}$，概率测度$P:\mathcal{F}\to[0,1]$。由三元组$(\Omega,\mathcal{F},P)$定义一个概率空间。

事件域$\mathcal{F}$是由样本空间的一些子集构成的集合，并满足以下条件：

- $\Omega\in\mathcal{F}$;
- 如果$A\in\mathcal{F}$, 则$\bar{A}\in\mathcal{F}$;
- 如果$A_n\in\mathcal{F}$, $n=1,\dots,\infty$, 则$\bigcup_{n=1}^{\infty}A_n\in\mathcal{F}$.

试证明：$\varnothing\in\mathcal{F}$. 如果$A_n\in\mathcal{F}$, $\bigcap_{n=1}^{\infty}A_n\in\mathcal{F}$.

概率的公理化定义

样本空间Ω，事件域$\mathcal{F}$，概率测度$P:\mathcal{F}\to[0,1]$。由三元组$(\Omega,\mathcal{F},P)$定义一个概率空间。

概率测度$P(A)$度量事件$A\in\mathcal{F}$发生的概率，满足下面三个公理。

- 公理1(非负性)：$\forall A\in\mathcal{F}$, $0\le P(A)$;
- 公理2(规范性)：$P(\Omega)=1$;
- 公理3(可列可加性)：对可列个两两互斥的事件$A_1,A_2,\dots,A_n,\dots$有，

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

思考：古典概率和几何概率是否符合该公理化定义?

概率的性质

1. $P(\varnothing)=0$
2. 如果A,B互斥，则$P(A+B)=P(A)+P(B)$.
3. 单调性：如果$A\subset B$,则$P(A)\le P(B)$.
4. $P(\bar{A})=1-P(A)$
5. $P(A\backslash B)=P(A)-P(AB)$
6. $P(A\cup B)=P(A)+P(B)-P(AB)$
7. 上连续性：如果$A_1\supset A_2\supset\cdots\supset A_n\cdots$，则
$$P\left(\bigcap_{i=1}^{\infty}A_i\right)=\lim_{i\to\infty}P(A_i).$$
8. 下连续性：如果$A_1\subset A_2\subset\cdots\subset A_n\cdots$，则
$$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\lim_{i\to\infty}P(A_i).$$

推论一

半可加性：对任意两个时间A, B, 有

$$P(A\cup B)\le P(A)+P(B).$$

对任意n个事件$A_1,A_2,\dots,A_n$, 有

$$P(\cup_{i=1}^{n}A_i)\le\sum_{i=1}^{n}P(A_i).$$

推论二

证明下列命题：

1. 若A_1与A_2同时发生时A发生，则有$P(A)\ge P(A_1)+P(A_2)-1$.
证明：因为$A_1A_2\subset A$，则$P(A)\ge P(A_1A_2)$. 又
$$\begin{aligned}P(A_1A_2)&=P(A_1)+P(A_2)-P(A_1\cup A_2)\\&\ge P(A_1)+P(A_2)-1.\end{aligned}\tag{1}$$
2. 若$A_1A_2A_3\subset A$，则有$P(A)\ge P(A_1)+P(A_2)+P(A_3)-2$.
证明：由第一题结论得到$P(A)\ge P(A_1A_2)+P(A_3)-1$. 再由(1)可得所需结果。

思考：若$\bigcap_{i=1}^{n}A_i\subset A$，则$P(A)\ge\sum_{i=1}^{n}P(A_i)-n+1$.

第一次作业

第一章习题：1.3, 1.4, 1.6, 1.12, 1.15, 1.18

第二章：一维随机变量及其分布

目录

为什么引进随机变量?

- 一些随机现象的样本点用数量表示，如骰子的点数、产品的寿命、测量误差
- 但有些随机现象的样本点本身不是数，如$\Omega = \{合格品, 不合格品\}$
- 随机变量将所有的样本点（基本事件）映射到实数域$\mathbb{R}$，这样便于定量化地分析和研究随机现象。

随机变量的概念

定义：考虑概率空间$(\Omega, \mathcal{F}, P)$, 称映射$X : \Omega \to \mathbb{R}$为随机变量，如果对任意$x \in \mathbb{R}$，有

$$\{\omega | X(\omega) \le x\} \in \mathcal{F}.$$

也就是$\{\omega | X(\omega) \le x\}$为一个随机事件。

注:

- 一般$X(\omega)$简记为X，$\{\omega | X(\omega) \le x\}$简记为$\{X \le x\}$
- X取值为有限或可列则称为离散随机变量
- X取值充满某个区间(a, b)则称为连续随机变量
- 对于同个概率空间可以构造多种随机变量
- 证明：$\{X > x\}, \{X \ge x\}, \{X < x\}, \{X = x\}$都为事件。

思考：满足什么条件的集合$A \subset \mathbb{R}$使得$\{X \in A\} \in \mathcal{F}$成立?

分布函数的定义

定义：设X是一个随机变量，对任意实数x，称

$$F(x) = P(X \le x)$$

为随机变量X的分布函数(cumulative distribution function, CDF)。为了强调与X的对应关系，有时写成$F_X(x)$。

注：分布函数将概率和函数建立起联系，由此可计算与随机变量X有关事件的概率。

分布函数的性质

- 单调性: 分布函数$F(x)$是$\mathbb{R}$上单调非减函数，即对任意$x_1 < x_2$, $F(x_1) \le F(x_2)$.
- 有界性: 对任意$x \in \mathbb{R}$, $F(x) \in [0, 1]$, 且

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

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

- 右连续性: 对任意$x_0 \in \mathbb{R}$,

$$F(x_0) = \lim_{x \to x_0^+} F(x) = F(x_0 + 0).$$

注：满足这三个基本性质成为判别某个函数是否成为分布函数的充要条件。

分布函数的性质

对任意$a < b, x \in \mathbb{R}$,

$$\begin{aligned} P(X = x) &= F(x) - F(x-0) \\ P(X > x) &= 1 - F(x) \\ P(X \geq x) &= 1 - F(x-0) \\ P(X < x) &= F(x-0) \\ P(a < X < b) &= F(b-0) - F(a) \\ P(a \leq X \leq b) &= F(b) - F(a-0) \\ P(a \leq X < b) &= F(b-0) - F(b-0) \\ P(a < X \leq b) &= F(b) - F(a) \end{aligned}$$

若$F(x)$在a与b处连续，则$F(a-0) = F(a)$, $F(b-0) = F(b)$

目录

一维离散型随机变量

定义：设离散型随机变量X的所有可能取值为$x_1, x_2, \cdots, x_n, \cdots$. 则称

$$p_i = p(x_i) = P(X = x_i), i = 1, 2, \ldots$$

为X的概率分布列（简称分布列），也可写成

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$	$\cdots$

或者写成

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots \end{pmatrix}$$

注意：分布列必须考虑所有的情况，即满足

- 非负性：对于所有的i, $p(x_i) \geq 0$
- 正则性：$\sum p(x_i) = 1$

离散随机变量的分布函数

由分布列不难求出离散随机变量X的分布函数

$$F(x) = \sum_{x_i \leq x} p(x_i).$$

该函数为右连续的阶梯函数。

例：设离散随机变量X的分布列为

X	-1	2	3
P	0.25	0.5	0.25

试求$P(X \leq 0.5), P(1.5 < X \leq 2.5)$，并写出$X$的分布函数。

解：

$$P(X \leq 0.5) = P(X = -1) = 0.25$$

$$P(1.5 < X \leq 2.5) = P(X = 2) = 0.5$$

$$F(x) = \begin{cases} 0, & x < -1, \\ 0.25, & -1 \leq x < 2, \\ 0.75, & 2 \leq x < 3, \\ 1, & x \geq 3 \end{cases}$$

常见的离散型分布

- 单点分布$P(X = c) = 1$，唯一取值。又称退化的分布。
- 二项分布$X \sim B(n, p)$，取值范围为$1, \ldots, n$
- 泊松分布$X \sim \mathrm{Pois}(\lambda)$，取值范围为非负整数
- 几何分布$X \sim \mathrm{Geo}(p)$，取值范围为正整数
- 超几何分布$h(n, N, M)$，取值范围$0, 1, \ldots, n$
- 负二项分布$Nb(r, p)$，取值范围$r, r+1, \cdots$

二项分布

定义：记X为n重伯努利试验中事件A出现的次数，$P(A) = p$，则有

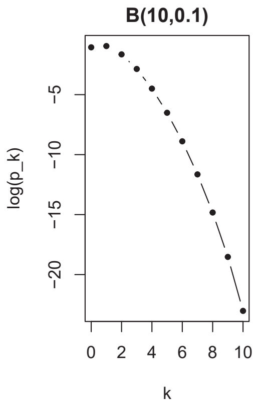
$$P(X = k) = b(k; n, p) = C_n^k p^k (1-p)^{n-k}, k = 0, \ldots, n.$$

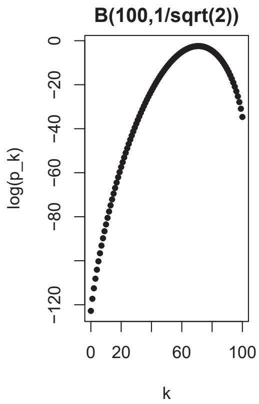
这样的分布称为二项分布，记为$X \sim B(n, p)$.

特别地，当$n = 1$时，二项分布也称为两点分布，其分布列为

$$\begin{pmatrix} 0 & 1 \\ p & 1-p \end{pmatrix}.$$

二项分布





二项分布

定理：令$m=(n+1)p$.

$$\arg\max_{k=0,\dots,n} b(k;n,p)=\begin{cases}\lfloor m\rfloor, \text{如果}m\text{不为整数}\\ \{m,m-1\}, \text{如果}m\text{为整数}\end{cases}$$

证明：

$$r=\frac{b(k;n,p)}{b(k-1;n,p)}=\frac{C_n^k p^k q^{n-k}}{C_n^{k-1}p^{k-1}q^{n-k+1}}=\frac{(n-k+1)p}{kq}=1+\frac{m-k}{kq}$$

如果$k<m$，则$r>1$，即$b(k;n,p)>b(k-1;n,p)$; 如果$k>m$, $b(k;n,p)<b(k-1;n,p)$. 所以当m不为整数时，$k=\lfloor m\rfloor$时，$b(k;n,p)$取得最大值。若m为整数，则$b(m;n,p)=b(m-1;n,p)$.

思考： 当k,n固定，p取何值时$b(k;n,p)$最大?

泊松分布

定义： 泊松分布是1837年由法国数学家泊松(Poisson, 1781–1840)首次提出的，泊松分布的概率分布列为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,\dots,$$

其中参数$\lambda>0$, 记为$X\sim \mathrm{Pois}(\lambda)$.

在实际问题中，有很多随机变量都近似服从泊松分布。例如:

- 在任给一段固定的时间间隔内，来到公共设施（公共汽车站、商店、电话交换台等）要求给予服务的顾客个数
- 炸弹爆炸后落在平面上某区域的碎弹片个数
- 落在显微镜片上的某种细菌个数

泊松定理——二项分布的泊松近似

定理： 设随机变量X服从二项分布$B(n,p_n)(n=1,2,\cdots)$，其中概率p_n与n有关，并且满足

$$\lim_{n\to\infty} np_n=\lambda>0,$$

则

$$\lim_{n\to\infty} P(X=k)=\lim_{n\to\infty} C_n^k p_n^k(1-p_n)^{n-k}=\frac{e^{-\lambda}\lambda^k}{k!}, k=0,1,2,\cdots.$$

证明： 令$\lambda_n=np_n$,则

$$\begin{aligned}b(k;n,p_n)&=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{n-k}\\&=\frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n-k}\\&\to\frac{e^{-\lambda}\lambda^k}{k!}.\end{aligned}$$

其中用到，$\lim_{n\to\infty}(1-\lambda_n/n)^{n-k}=\lim_{n\to\infty}[(1-\lambda_n/n)^{-n/\lambda_n}]^{-\lambda_n}=e^{-\lambda}$.

泊松定理的应用

在应用中，当n很大($n\geq 10$)，p很小($p\leq 0.1$)时，可以用以下近似

$$b(k;n,p)\approx\frac{e^{-\lambda}\lambda^k}{k!}, \lambda=np.$$

例： 设每次击中目标的概率为0.001，且各次射击是否中目标可看作相互没有影响，如果射击5000次，试求：（1）击中12弹的概率；（2）至少击中12弹的概率。

解： 设击中次数为X,则$X\sim B(5000,0.001)$. $\lambda=5000\times 0.001=5$.

(1)

$$P(X=12)\approx\frac{e^{-\lambda}\lambda^{12}}{12!}=\frac{e^{-5}5^{12}}{12!}=0.00343424(0.00342154).$$

(2)

$$P(X<12)\approx\sum_{k=0}^{11}\frac{e^{-5}5^k}{k!}=0.9945469.$$

则有$P(X\geq 12)\approx 0.0054531(0.0054284)$.

例

设有同类设备80台，各台工作相互独立的，发生故障的概率都是0.01，并且一台设备的故障可由一个人来处理，试求

- 一个人负责维修20台设备时，设备发生故障而不能及时维修的概率；
- 由三个人共同负责维修80台设备时，设备发生故障而不能及时维修的概率。

解： (1)设同一时刻发生故障的数目为X. $X\sim B(20,0.01)$. 求得

$$P(X\geq 2)\approx 0.0176$$

(2)设同一时刻发生故障的数目为Y. $Y\sim B(80,0.01)$. 求得

$$P(Y\geq 4)\approx 0.0091$$

计算结果表明，由三人共同负责维修80台，每人平均约维修27台，比一个单独维修20台更好，既节约了人力又提高了工作效率。

几何分布

定义：在"成功"概率是p的伯努利试验中，若以X记首次出现"成功"的试验次数。则X所服从的分布便是几何分布，分布列为

$$P(X=k)=g(k;p)=(1-p)^{k-1}p, k=1,2,\cdots.$$

记为$X\sim \mathrm{Geo}(p)$.

例：一个人要开门，他共有n把钥匙，其中仅有一把是能开此门的，现随机地从中取出一把钥匙来试开门，在试开时每一把钥匙均以$1/n$的概率被取用，问此人直到第k次试开时方才成功的概率是多少?

$$g(k;1/n)=\left(1-\frac{1}{n}\right)^{k-1}\frac{1}{n}.$$

几何分布的特征——无记忆性

定理：设$X\sim \mathrm{Geo}(p)$, 则有对于任意正整数k,n,

$$P(X=k+n|X>k)=P(X=n).$$

注：已知第k次还未成功，那么从第$k+1$次开始，首次成功出现在哪一次与k无关。

证明：

$$\begin{aligned}P(X=k+n|X>k)&=\frac{P(X=k+n)}{P(X>k)}\\&=\frac{(1-p)^{k+n-1}p}{\sum_{i=k+1}^{\infty}(1-p)^{i-1}p}\\&=\frac{(1-p)^{k+n-1}p}{(1-p)^k}\\&=(1-p)^{n-1}p=P(X=n).\end{aligned}$$

推论：$P(X>k+n|X>k)=P(X>n)$.

超几何分布

定义：在一箱N件装的产品中混进了M件次品，今从中抽取n件($n\le M$)，求从中查出次品的件数X的概率分布.

$$P(X=k)=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}, k=0,1,\dots,n.$$

记为$X\sim h(n,N,M)$.

当$n\ll N$, 即抽取个数n远小于总数N,每次抽取后总体的不合格率$p=M/N$改变甚微，所以可以视为不放回抽样，这是超几何分布可以用二项分布近似：

$$P(X=k)\approx b(k;n,M/N)$$

负二项分布(帕斯卡分布)

定义：在"成功"概率是p的伯努利试验中，出现第r次成功时所作的试验次数X所服从的分布称为负二项分布，分布列为

$$P(X=k)=f(k;r,p)=C_{k-1}^{r-1}p^r(1-p)^{k-r}, k=r,r+1,\cdots$$

记为$X\sim Nb(r,p)$.

由于$f(k;r,p)$是负指数二项式$(1/p-q/p)^{-r}$展开式中的项($q=1-p$)，故X所服从的分布称为负二项分布。由此也可以证明

$$\begin{aligned}\sum_{k=r}^{\infty}f(k;r,p)&=p^r\sum_{k=r}^{\infty}C_{k-1}^{r-1}q^{k-r}=p^r\sum_{k=0}^{\infty}C_{r+k-1}^{r-1}q^k\\&=p^r\sum_{k=0}^{\infty}C_{-r}^k(-q)^k=p^r(1-q)^{-r}=1.\end{aligned}$$

注：$C_{-r}^k=(-r)(-r-1)\dots(-r-k+1)/k!=(-1)^rC_{r+k-1}^k=(-1)^rC_{r+k-1}^{r-1}$.

几何分布与负二项分布的关系

若$X_i,i=1,\dots,r$独立同分布，且$X_i\sim \mathrm{Geo}(p)$. 则

$$X_1+X_2+\cdots+X_r\sim Nb(r,p).$$

即负二项分布随机变量可以分解成r个独立同分布的几何分布随机变量之和。

目录

一维连续型随机变量

定义：设随机变量X的分布函数为$F(x)$，若存在非负函数$f(x)$，使得对一切实数x，恒有

$$F(x) = \int_{\infty}^{x} f(t)\,\mathrm{d}t,$$

则称X为连续型随机变量，$f(x)$称为X的密度函数(Probability Distribution Function, PDF)。

密度函数的最基本性质：

- 非负性

$$f(x) \geq 0$$

- 正则性

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

连续型随机变量的性质

1. 连续型随机变量的分布函数是连续函数
2. $$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)\,\mathrm{d}x$$
3. 若$f(x)$在x的邻域连续，则$f(x) = F'(x)$
4. 对于任意实数a, $P(X = a) = 0$
5. $$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b)$$

例1

设随机变量X的分布函数为：

$$F(x) = \begin{cases} A + \frac{1}{3}e^{x}, & x < 0 \\ B - \frac{1}{3}e^{-2x}, & x \geq 0 \end{cases}$$

1. 求常数A, B
2. 判断X是否为连续型随机变量
3. 求$P(-1 < X \leq 1/2)$

证明：$F(-\infty) = A = 0$，$F(\infty) = B = 1$. 因为$F(0) = 2/3 \neq F(0-) = 1/3$，所以$F(x)$在$x = 0$处不连续，也就意味着$X$不为连续型随机变量。

$$P(-1 < X \leq 1/2) = F(1/2) - F(-1) = 1 - \frac{2}{3e}$$

注：X既不是离散型也不是连续型随机变量。

例2

设随机变量X的密度函数为：

$$f(x) = \frac{a}{1+x^2}, x \in (-\infty, \infty).$$

1. 试确定a的值
2. 试求X的分布函数
3. 试求$P(X^2 \leq 1)$

证明：由正则性得

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

所以$a = 1/\pi$.

$$F(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+t^2}\,\mathrm{d}t = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$$

$$P(X^2 \leq 1) = F(1) - F(-1) = [\arctan(1) - \arctan(-1)]/\pi = 0.5$$

注：此分布为柯西(Cauchy)分布

常见的连续型分布

1. 均匀分布
2. 指数分布
3. 正态分布
4. 卡方分布，t分布，F分布(pp.104-106)
5. 伽玛分布
6. 贝塔分布
7. 对数正态分布
8. 柯西分布

1. 均匀分布

定义：设a、b为有限数，且$a < b$。如果随机变量X分布密度为

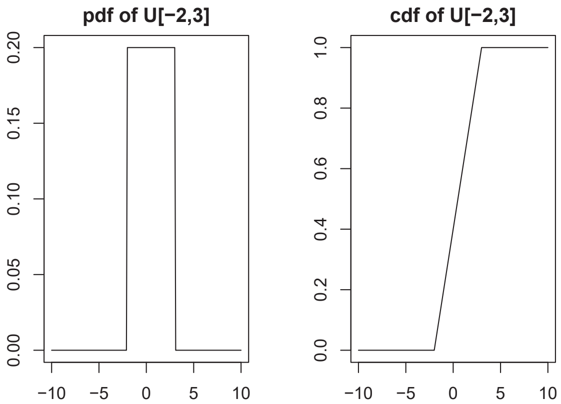
$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

则称X服从$[a, b]$上的均匀分布，记$X \sim U[a, b]$. 其分布函数为

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

注：$U[0, 1]$是最基本的随机变量，由此可以生成任意的分布。

均匀分布密度与分布函数示意图



例

向区间$(-1,1)$均匀地投掷一随机点，以X表示随机点的落点坐标，试求关于t的二次方程$t^2+3Xt+1=0$有实根的概率。

解：该二次方程有实根的充要条件为$9X^2-4\geq 0$. 依题意$X\sim U(-1,1)$. 则

$$P(9X^2-4\geq 0)=P(|X|\geq 2/3)=\int_{|x|\geq 2/3}\frac{1}{2}\mathbf{1}\{-1<x<1\}\,\mathrm{d}x$$
$$=\int_{-1}^{-\frac{2}{3}}\frac{1}{2}\,\mathrm{d}x+\int_{\frac{2}{3}}^{1}\frac{1}{2}\,\mathrm{d}x=\frac{1}{3}.$$

2. 指数分布

定义：若随机变量X具有分布密度为

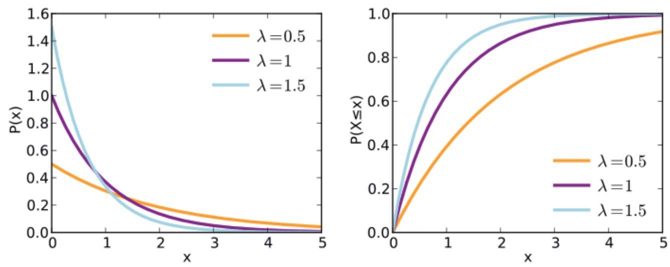
$$f(x)=\begin{cases}\lambda\exp(-\lambda x), & x\geq 0\\ 0, & x<0.\end{cases}$$

其中$\lambda>0$，则称X服从指数分布，记为$X\sim Exp(\lambda)$. 其分布函数为

$$F(x)=\begin{cases}1-\exp(-\lambda x), & x\geq 0\\ 0, & x<0.\end{cases}$$

注：指数分布常被用作各种"寿命"分布，譬如电子元件的寿命，动物的寿命，电话的通话时间，随机服务系统的服务时间。在可靠性与排队论中广泛应用。

指数分布的密度与分布函数示意图



指数分布的无记忆性

定理：如果$X\sim Exp(\lambda)$, 则对任意$s,t>0$,有

$$P(X>s+t|X>s)=P(X>t)$$

证明：

$$P(X>s+t|X>s)=\frac{P(X>s+t,X>s)}{P(X>s)}=\frac{P(X>s+t)}{P(X>s)}$$
$$=\frac{\exp(-\lambda(s+t))}{\exp(-\lambda(s))}=\exp(-\lambda t)=P(X>t)$$

注：还可以证明指数分布是连续型随机变量中唯一的具有无记忆性的分布。

例

设到某服务窗口办事，需排队等候。若等待的时间X服从参数为$1/10$的指数分布随机变量（单位:分钟），若等待超过15分钟后仍未能得到接待时，他就愤然离去。试求:

- 10位顾客有2位愤然离去的概率;
- 10位顾客最多有2位愤然离去的概率;
- 10位顾客至少有2位愤然离去的概率。

解：顾客愤然离去的概率

$$p=P(X>15)=1-F(15)=\exp(-15/10)=0.2231$$

设10位顾客离开的个数为Y, 则$Y\sim B(10,0.2231)$. (1)$P(Y=2)=0.2937$;(2)$P(Y\leq 2)=0.6735$;(3)$P(Y\geq 2)=1-P(Y<2)=0.6238$.

3. 正态分布

定义：若随机变量X的分布密度

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

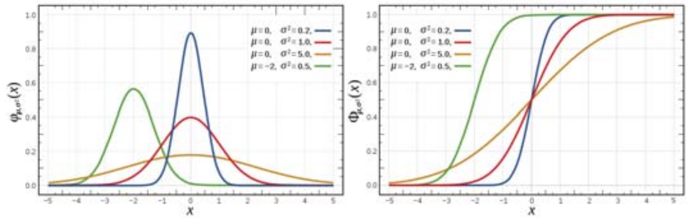
其中$\mu\in\mathbb{R},\sigma>0$, 则称$X$服从参数为$\mu,\sigma$的正态分布，简记为$X\sim N(\mu,\sigma^2)$。其分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t.$$

特别地，$\mu=0,\ \sigma=1$, $N(0,1)$称标准正态分布，其密度和分布函数常用ϕ,Φ表示。

注：正态分布的分布函数没有解析解，可通过查表得到，常用的数学软件使用的是它的近似值。很多随机变量可以用正态分布来描述或者近似描述，譬如测量误差、产品重量、人的身高、年降雨量。

正态分布的密度与分布函数示意图



- 密度函数关于$X=\mu$对称，μ称为位置参数
- σ称为尺度参数，当μ不变时，σ越小，曲线呈现高瘦，σ越大，曲线呈现矮胖。

标准正态分布

标准正态分布$X\sim N(0,1)$, 密度函数

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

分布函数

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}\,\mathrm{d}t$$

- $\phi(x)=\phi(-x)$
- $\Phi(-x)=1-\Phi(x)$, 查表只给出分布函数在$x\geq 0$的近似值，由此可得$x<0$时$\Phi(x)$的值
- $P(|X|\leq c)=2\Phi(c)-1$，其中$c\geq 0$.
- $P(|X|\geq c)=2(1-\Phi(c))$，其中$c\geq 0$.
- $P(|X|\leq 1)\approx 0.6827$, $P(|X|\leq 2)\approx 0.9545$, $P(|X|\leq 3)\approx 0.9973$

一般正态分布的性质

定理：若$X\sim N(\mu,\sigma^2)$, 则

$$\frac{X-\mu}{\sigma}\sim N(0,1).$$

由此可得：

- $f(x+\mu)=f(\mu-x)$
- $F(x)=\Phi((x-\mu)/\sigma)$
- $F(\mu-x)=1-F(\mu+x)$
- $P(|X-\mu|\leq\sigma)\approx 0.6827$
- $P(|X-\mu|\leq 2\sigma)\approx 0.9545$
- $P(|X-\mu|\leq 3\sigma)\approx 0.9973$, 说明$X$的99.73%的值落在$(\mu-3\sigma,\mu+3\sigma)$内。这个性质被实际工作者称为正态分布的“$3\sigma$原则”。

例1

设$X\sim N(-1,4)$, 试求$P(-5\leq X<1)$, $P(|X|<1)$, $P(|X|\geq 3/2)$.

解：

$$\begin{aligned}P(-5\leq X<1)&=\Phi((1-(-1))/2)-\Phi((-5-(-1))/2)=\Phi(1)-\Phi(-2)\\&=\Phi(1)-(1-\Phi(2))=\Phi(1)+\Phi(2)-1=0.8186.\end{aligned}$$

$$P(|X|<1)=\Phi((1-(-1))/2)-\Phi((-1-(-1))/2)=\Phi(1)-\Phi(0)=0.3413$$

$$\begin{aligned}P(|X|\geq 3/2)&=1-P(|X|<3/2)=1-\Phi(5/4)+\Phi(-1/4)\\&=2-\Phi(5/4)-\Phi(1/4)=0.5070\end{aligned}$$

例2

公共汽车车门的高度是按男子与车门顶碰头的机会在0.01以下来设计的，设男子身高X服从$\mu=170cm,\sigma=6cm$的正态分布，即$X\sim N(170,6^2)$，试确定车门的高度。

解：设车门的高度为h, 依题意有

$$P(X>h)<0.01$$

则

$$P(X\leq h)=\Phi((h-170)/6)>0.99$$

即

$$\frac{h-170}{6}\geq\Phi^{-1}(0.99)=2.33$$

所以

$$h\geq 170+6\times 2.33=184$$

例3：股价变化幅度的估计

设某只股票的初始价格为$S_0 = 40$元，预期年收益率$\mu = 16\%$，年波动率为$\sigma = 20\%$. 在Black-Scholes模型下（1997年诺贝尔经济学奖得主），股票在每个时刻t的价格S_t为随机变量，且

$$S_t = S_0 \exp\{(\mu - \sigma^2/2)t + \sigma B_t\},$$

其中$B_t \sim N(0, t)$. 试估计六个月后这只股票的价格范围（允许出错的概率为5%）

解：令$Y_t = [\log(S_t) - (\mu - \sigma^2/2)t]/(\sigma\sqrt{t}) \sim N(0, 1)$。求$y$使得

$$P(|Y_t| \le y) = 0.95$$

因为$P(|Y_t| \le y) = 2\Phi(y) - 1 = 0.95$, 所以$\Phi(y) = 0.975$, $y = \Phi^{-1}(0.975) = 1.96$
令$t = 1/2$, 由$P(|Y_{1/2}| \le 1.96) = 0.95$得到$P(32.51 \le S_{1/2} \le 56.6) = 0.95$.

目录

一维随机变量函数的分布

已知随机变量X的分布，如何求另一个随机变量$g(X)$的分布？先考虑离散的情形。

设离散型随机变量X的分布列为：

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$	$\cdots$

则$Y = f(X)$也为离散型随机变量，其分布列为

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_n)$	$\cdots$
P	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$	$\cdots$

注：当$g(x_1), g(x_2), \cdots, g(x_n), \cdots$中有某些值相等时，则把相等的值分别合并。

例

设X的分布列为：

X	-2	-1	0	1	2
P	0.15	0.2	0.2	0.2	0.25

求$Y = 2X - 1, Z = X^2$的分布列。

解：

$Y = 2X - 1$	-5	-3	-1	1	3
P	0.15	0.2	0.2	0.2	0.25

$Z = X^2$	4	1	0	1	4
P	0.15	0.2	0.2	0.2	0.25

$Z = X^2$	0	1	4
P	0.2	0.4	0.4

连续随机变量函数的分布

定理：设X为连续随机变量，其密度函数为$p_X(x)$. $Y = g(X)$是另一个随机变量。若$y = g(x)$严格单调，其反函数$h(y)$有连续的导函数，则$Y = g(X)$的密度函数为

$$f_Y(y) = \begin{cases} f_X(h(y))\,|h'(y)|, & y \in (a, b) \\ 0, & \text{其他}, \end{cases} \tag{1}$$

其中$a = \min\{g(\infty), g(-\infty)\}, b = \max\{g(\infty), g(-\infty)\}$.

证明：不妨设g严格单调递增。

$$F_Y(y) = P(g(X) \le y) = P(X \le h(y)) = F_X(h(y)).$$

$$f_Y(y) = F_Y'(y) = \frac{\mathrm{d}F_X(h(y))}{\mathrm{d}y} = f_X(h(y))h'(y)$$

$g(x)$为线性函数

特别地，当$g(x) = ax + b$, $a \ne 0$, 则$Y = g(X)$的密度函数为

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right).$$

定理：设$X \sim N(\mu, \sigma^2)$。则

- $Y = (X - \mu)/\sigma \sim N(0, 1)$
- $Z = aX + b \sim N(a\mu + b, a^2\sigma^2)$, 其中$a \ne 0$.

证明：(1)$a = 1/\sigma$, $b = -\mu/\sigma$,

$$f_Y(y) = \sigma f_X((\sigma y + \mu)) = \sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\sigma y + \mu - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

(2)由(1)知，$X = \mu + \sigma X_0$, 其中$X_0 \sim N(0, 1)$. 所以，

$$Z = aX + b = a(\mu + \sigma X_0) + b = a\mu + b + a\sigma X_0 \sim N(a\mu + b, a^2\sigma^2)$$

$g(x)$为分布函数

定理：若随机变量X的分布函数$F_X(x)$为严格单调递增的连续函数，其反函数F_X^{-1}存在，则$Y = F_X(X) \sim \mathbb{U}(0,1)$.

证明： 当$y \le 0$时，$P(Y \le y) = 0$; 当$y \ge 1$时，$P(Y \le y) = 1$.
当$y \in (0,1)$时，

$$P(Y \le y) = P(F_X(X) \le y) = P(X \le F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y.$$

注：已知F_X^{-1}，则可以通过变换$X = F_X^{-1}(U)$来生成X的样本，其中，$U \sim \mathbb{U}(0,1)$。这种方式称为逆变换方法。由此可以看出，任何分布可以由均匀分布（在计算机中为伪随机数，如，Mersenne–Twister算法）变换产生。

一般的$g(x)$

一般情况下$Y = g(X)$的密度没有通用的表达式，通常的做法是按定义求分布函数，转化成计算以下概率：

$$F_Y(y) = P(g(X) \le y).$$

然后再求导得到密度。

例：如果$X \sim N(0,1)$，试求$Y = X^2$的密度函数。

解：

$$F_Y(y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1$$

$$f_Y(y) = F_Y'(y) = \frac{\phi(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, y > 0.$$

第三次作业

第二章习题：4, 7, 9, 13, 16, 19, 21

第三章：随机向量及其分布

目录

随机向量的概念

定义：设$X_1, \ldots, X_n$为概率空间$(\Omega, \mathcal{F}, P)$的n个随机变量，则称向量

$$\mathbf{X} := (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$$

为n维随机向量。

注：多维随机向量的关键是定义在同个概率空间上。对于不同样本空间Ω_1, Ω_2上的两个随机变量，我们在乘积空间

$$\Omega = \Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2) | \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

及其事件域上讨论。

随机向量的分布函数

定义：设$\mathbf{X} = (X_1, \ldots, X_n)$为概率空间$(\Omega, \mathcal{F}, P)$的随机向量，则它的联合分布函数定义为:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(\cap_{i=1}^n \{X_i \le x_i\}) = P\left(\mathbf{X} \in \prod_{i=1}^n (-\infty, x_i]\right).$$

随机向量分布函数的特征性质：

1. $0 \le F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \le 1, \ \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
2. $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$关于每个变量$x_i$单增右连续
3. $\lim_{x_i \to -\infty} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = 0$, $\lim_{\min x_i \to \infty} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = 1$
4. $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$, $\forall h_i > 0$,

$$\Delta_{(x_1,\ldots,x_n)}^{(x_1+h_1,\ldots,x_n+h_n)} F_{X_1,\ldots,X_n}(t_1, \ldots, t_n) \ge 0.$$

注：上式n阶差分等价于$P(\cap_{i=1}^n \{x_i < X_i \le x_i + h_i\})$.

边缘分布

已知随机向量$\mathbf{X} = (X_1, \ldots, X_n)$的联合分布为$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$, 由此可确定每个分量的边缘分布$F_{X_i}$,

$$F_{X_i}(x_i) = \lim_{x_j \to \infty, j \ne i} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n).$$

更一般地，令$A = \{i_1, \ldots, i_k\} \subseteq I = \{1, \ldots, n\}$,

$$F_{X_{i_1},\ldots,X_{i_k}}(x_{i_1}, \ldots, x_{i_k}) = \lim_{x_j \to \infty, j \in I \setminus A} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n).$$

离散型与连续型

定义：设$\mathbf{X} = (X_1, \ldots, X_n)$为概率空间$(\Omega, \mathcal{F}, P)$的随机向量，

- 如果$(X_1, \ldots, X_n)$至多取可数个不同的值，则称之为离散型随机向量;
- 如果存在非负函数$f_{X_1,\ldots,X_n}$使得$(X_1, \ldots, X_n)$的联合分布函数可以表示为

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \mathrm{d}t_1 \int_{-\infty}^{x_2} \mathrm{d}t_2 \cdots \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(t_1, t_2, \ldots, t_n)\, \mathrm{d}t_n,$$

  则称$(X_1, \ldots, X_n)$为连续型随机向量，称$f_{X_1,\ldots,X_n}$为它的分布密度函数。对于任意区域$D \subset \mathbb{R}^n$, 有

$$P(\mathbf{X} \in D) = \int \cdots \int_D f_{X_1,\ldots,X_n}(t_1, \ldots, t_n)\, \mathrm{d}t_1 \cdots \mathrm{d}t_n.$$

随机变量的独立性

定义：设$\mathbf{X} = (X_1, \ldots, X_n)$为概率空间$(\Omega, \mathcal{F}, P)$的随机向量，如果

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n), \ \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

则称$X_1, \ldots, X_n$相互独立。

等价形式：任意集合$A_1, \ldots, A_n$, 满足

$$P(\cap_{i=1}^n \{X_i \in A_i\}) = \Pi_{i=1}^n P(X_i \in A_i).$$

即所有与X_i相关的事件都相互独立。

独立性判别方法

定理：设$\mathbf{X} = (X_1, \ldots, X_n)$为概率空间$(\Omega, \mathcal{F}, P)$的随机向量，

- 如果$X_1, \ldots, X_n$都为离散型随机变量，有分布列

$$P(X_i = a_j^{(i)}), j = 1, 2, \cdot, i = 1, \ldots, n,$$

  则$X_1, \ldots, X_n$相互独立的充要条件为

$$P(X_1 = a_{\ell_1}^{(1)}, X_2 = a_{\ell_2}^{(2)}, \ldots, X_n = a_{\ell_n}^{(n)}) = P(X_1 = a_{\ell_1}^{(1)}) \cdots P(X_n = a_{\ell_n}^{(n)}),$$

  其中ℓ_i为任意正整数。
- 如果$X_1, \ldots, X_n$都为连续型随机变量，有联合分布密度函数$f_{X_1,\ldots,X_n}$，则$X_1, \ldots, X_n$相互独立的充要条件为

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

目录

二维离散型随机变量

定义： 设二维离散型随机变量(X, Y)的取值为$(x_i, y_j), i, j = 1, 2\cdots$. 分布列为

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

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot} = \sum_j p_{1j}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot} = \sum_j p_{2j}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot} = \sum_j p_{ij}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$p_{\cdot j}$	$p_{\cdot 1} = \sum_i p_{\cdot 1}$	$p_{\cdot 2} = \sum_i p_{i2}$	$\cdots$	$p_{\cdot j} = \sum_i p_{ij}$	$\cdots$	1

其中，X的边缘分布为$P(X = x_i) = p_{i\cdot}$，Y的边缘分布为$P(Y = y_j) = p_{\cdot j}$.

例1

例： 袋中有五件产品，其中两件次品，三件正品，从袋中任意依次取出两件，每次取出的产品进行检查后放回袋中，设每次取出产品时，袋中每件产品被取到的可能性相等，求下列随机变量的联合分布列以及边缘分布：

$$X = \begin{cases} 1, & \text{第一次取到正品} \\ 0, & \text{第一次取到次品} \end{cases}, \quad Y = \begin{cases} 1, & \text{第二次取到正品} \\ 0, & \text{第二次取到次品} \end{cases}$$

解：

$X\backslash Y$	0	1	
0	$\frac{4}{25}$	$\frac{6}{25}$	$\frac{2}{5}$
1	$\frac{6}{25}$	$\frac{9}{25}$	$\frac{3}{5}$
	$\frac{2}{5}$	$\frac{3}{5}$	1

例2

例： 在例 1 中，如果每次取出后不放回，求(X, Y)的联合分布列以及边际分布。

解：

$X\backslash Y$	0	1	
0	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{5}$
1	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{3}{5}$
	$\frac{2}{5}$	$\frac{3}{5}$	1

例3

设随机试验只有A、B和C三个结果，各结果出现的概率分别是p, q和$1 - p - q$. 现将该随机试验独立重复做n次，记X和Y分别为n次试验中A和B发生的次数，试求(X, Y)的联合分布和边缘分布。

解：

$$P(X = i, Y = j) = C_n^i C_{n-i}^j p^i q^j (1 - p - q)^{n-i-j}, \ 0 \le i + j \le n.$$

边缘分布$X \sim B(n, p)$, $Y \sim B(n, q)$.

二维离散型随机向量条件分布列

独立： X, Y相互独立的充要条件是对所有可能的取值，有

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i),$$

即$p_{ij} = p_{i\cdot} p_{\cdot j}$, $i, j \ge 1$.
注： 例1独立，例2和例3不独立。
条件分布列： 设(X, Y)为二维离散型随机变量，其联合分布列为

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

已知事件$\{Y = y_j\}$发生，在此条件下X的分布列称条件分布列，

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

同样可以定义已知事件$\{X = x_i\}$发生，在此条件下Y的条件分布列。独立情况下，条件分布退化成无条件分布。

例：

考虑例2，求$\{X = 1\}$条件下Y的条件分布列，$\{Y = 0\}$条件下X的条件分布列。

$X\backslash Y$	0	1	
0	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{5}$
1	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{3}{5}$
	$\frac{2}{5}$	$\frac{3}{5}$	1

解：

$$P(Y = 0|X = 1) = P(Y = 1|X = 1) = \frac{3/10}{3/5} = \frac{1}{2}$$

$$P(X = 0|Y = 0) = \frac{1/10}{2/5} = \frac{1}{4}, \ P(X = 1|Y = 0) = \frac{3/10}{2/5} = \frac{3}{4}.$$

目录

二维连续型随机向量

定义：若(X, Y)为连续型随机型随机向量，则其分布函数

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\,\mathrm{d}u\,\mathrm{d}v.$$

分布密度函数$f_{X,Y}$满足：

- 非负性：$f_{X,Y}(x, y) \geq 0,\ (x, y) \in \mathbb{R}^2$
- 正则性：$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\,\mathrm{d}u\,\mathrm{d}v = 1$
- 若$f_{X,Y}(x, y)$在点(x_0, y_0)处连续，则
$$f_{X,Y}(x_0, y_0) = \frac{\partial^2 F_{X,Y}(x_0, y_0)}{\partial x \partial y}$$
- 对二维平面的任何区域D有
$$P((X, Y) \in D) = \iint_D f_{X,Y}(u, v)\,\mathrm{d}u\,\mathrm{d}v.$$

二维连续型随机向量的边缘分布

设(X, Y)为连续型随机型随机向量，其密度函数为$f_{X,Y}(x, y)$，则X的边缘分布函数、密度函数分别为

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\,\mathrm{d}v\,\mathrm{d}u.$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y.$$

同样，Y的边缘密度函数为

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}x.$$

二维均匀分布

定义：设D为二维平面上的一个有界区域，面积为$S_D < \infty$，若随机向量(X, Y)的分布密度函数为

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{S_D}, & (x, y) \in D, \\ 0, & \text{其他}, \end{cases}$$

则称(X, Y)服从D上的均匀分布。

二维正态分布

定义：若随机向量(X, Y)的分布密度函数为

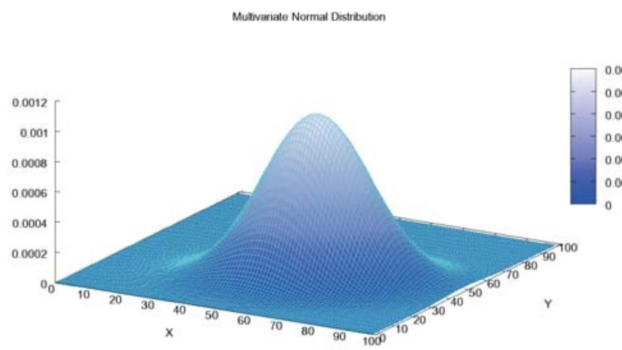
$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\},$$

其中$\sigma_1, \sigma_2 > 0, |\rho| < 1$，则称$(X, Y)$服从参数为$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$的正态分布，记为

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

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

二维正态分布的密度函数



n维正态分布

定义：更一般地，n维正态分布$\mathbf{X}=(X_1,\dots,X_n)^\top$的密度函数为

$$f_{\mathbf{X}}(x_1,\dots,x_n)=\frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\{-(1/2)(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$$

其中$\boldsymbol{\mu}=(\mu_1,\dots,\mu_n)^\top$，$\boldsymbol{\Sigma}$为$n$阶正定矩阵，记$X\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

注1：$X_i\sim N(\mu_i,\Sigma_{ii})$，$i=1,\dots,n$.

注2：$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$等价于$(X,Y)\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$，其中$\boldsymbol{\mu}=(\mu_1,\mu_2)^\top$，

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}.$$

例1

例：设二维连续型随机变量(X,Y)的分布函数为

$$F(x,y)=(A+B\arctan x)(C+\arctan y).$$

- 求常数A,B,C;
- (X,Y)的密度函数;
- $D=\{(x,y)|x-y>0,x\le 1\}$，求$P\{(X,Y)\in D\}$.

解：(1)由二维分布函数的性质知，

$$F(-\infty,y)=(A-\frac{\pi}{2}B)(C+\arctan y)=0$$
$$F(x,-\infty)=(A+B\arctan x)(C-\frac{\pi}{2})=0$$
$$F(\infty,\infty)=(A+\frac{\pi}{2}B)(C+\frac{\pi}{2})=1,$$

可得$C=\pi/2$，$A=1/(2\pi)$，$B=1/\pi^2$.

例1

例：设二维连续型随机变量(X,Y)的分布函数为

$$F(x,y)=(A+B\arctan x)(C+\arctan y).$$

- (X,Y)的密度函数;
- $D=\{(x,y)|x-y>0,x\le 1\}$，求$P\{(X,Y)\in D\}$.

解：(2)密度函数

$$f(x,y)=\frac{\partial^2F(x,y)}{\partial x\partial y}=\frac{1}{\pi^2(1+x^2)(1+y^2)}$$

(3)

$$P\{(X,Y)\in D\}=\int_{(x,y)\in D}f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\frac{1}{\pi^2}\int_{-\infty}^{1}\frac{1}{1+x^2}\int_{-\infty}^{x}\frac{1}{1+y^2}\,\mathrm{d}y\,\mathrm{d}x$$
$$=\frac{1}{\pi^2}\left[\frac{1}{2}(\arctan x)^2+\frac{\pi}{2}\arctan x\right]_{-\infty}^{1}=\frac{9}{32}.$$

例2

例：已知二维随机向量(X,Y)的密度为

$$f(x,y)=\begin{cases}kxy & x^2\le y\le 1,0\le x\le 1\\ 0 & \text{其他}\end{cases}$$

试确定k的数值，并求(X,Y)落在区域$D=\{(x,y)|x^2\le y\le x,0\le x\le 1\}$的概率。

解：(1)由概率密度性质，知

$$\int\int f(x,y)\,\mathrm{d}x\,\mathrm{d}y=k\int_0^1 x\int_{x^2}^1 y\,\mathrm{d}y\,\mathrm{d}x=\frac{k}{6}=1\Rightarrow k=6.$$

(2)

$$P\{(X,Y)\in D\}=6\int_0^1 x\int_{x^2}^{x}y\,\mathrm{d}y\,\mathrm{d}x=\frac{1}{4}.$$

二维连续型随机向量的独立性

独立的充要条件：设(X,Y)是二维连续型随向量，$f(x,y)$及$f_X(x),f_Y(y)$分别是(X,Y)的联合分布密度及边缘分布密度，则X,Y相互独立的充要条件是：对任意点(x,y)，有

$$f(x,y)=f_X(x)f_Y(y).$$

例：易知例1独立。现考虑例2，当$0\le x,y\le 1$时，

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y=6x\int_{x^2}^{1}y\,\mathrm{d}y=3x-3x^5.$$

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x=6y\int_0^{\sqrt{y}}x\,\mathrm{d}x=3y^2.$$

因为$f(x,y)\ne f_X(x)f_Y(y)$，所以X,Y不独立。

二维正态分布独立的充要条件

定理：若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，$X$与$Y$独立的充要条件是$\rho=0$.

注：ρ称为相关系数，刻画X与Y的相关性。

证明：若X,Y独立，则有$f(x,y)=f_X(x)f_Y(y)$. 从而有

$$\sup_{x,y}f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}=\sup_{x,y}f_X(x)f_Y(y)=\frac{1}{2\pi\sigma_1\sigma_2},$$

所以$\rho=0$. 反过来，如果$\rho=0$，则易知，$f(x,y)=f_X(x)f_Y(y)$.

二维连续型随机向量的条件分布

设(X,Y)是二维连续型随机向量，先考虑$Y=y$的条件下，X的条件分布。

$$
\begin{aligned}
P(X\le x|Y=y) &= \lim_{\Delta y\to 0+} P(X\le x|y\le Y\le y+\Delta y)\\
&= \lim_{\Delta y\to 0+}\frac{P(X\le x, y\le Y\le y+\Delta y)}{P(y\le Y\le y+\Delta y)}\\
&= \lim_{\Delta y\to 0+}\frac{\int_{\infty}^{x}\int_{y}^{y+\Delta y} f_{X,Y}(u,v)\,\mathrm{d}v\,\mathrm{d}u}{\int_{y}^{y+\Delta y} f_Y(v)\,\mathrm{d}v} \quad \text{中值定理}\\
&= \lim_{\Delta y\to 0+}\frac{\int_{\infty}^{x}\Delta y f_{X,Y}(u,v^*)\,\mathrm{d}u}{\Delta y f_Y(v^{**})} = \int_{\infty}^{x}\frac{f_{X,Y}(u,y)}{f_Y(y)}\,\mathrm{d}u
\end{aligned}
$$

定义：如果$f_Y(y)>0$, 在$Y=y$的条件下, X的条件密度可以定义为:

$$f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)},\ \forall x\in\mathbb{R}.$$

如果$f_X(x)>0$，同样方式可以定义$f_{Y|X}(y|x)$. 独立情况下，条件分布退化成无条件分布。

例2（续）

例：已知二维随机向量(X,Y)的密度为

$$f(x,y)=\begin{cases}6xy & x^2\le y\le 1, 0\le x\le 1\\ 0 & \text{其他}\end{cases}$$

求$f_{X|Y}(x|y)$和$f_{Y|X}(y|x)$.

解：已知

$$f_X(x)=\begin{cases}3x-3x^5 & 0\le x\le 1\\ 0 & \text{其他}\end{cases},\quad f_Y(y)=\begin{cases}3y^2 & 0\le y\le 1\\ 0 & \text{其他}\end{cases}$$

所以，当$0<y\le 1$, 当$0<x<1$时,

$$f_{X|Y}(x|y)=\begin{cases}2x/y & 0\le x\le\sqrt{y}\\ 0 & \text{其他}\end{cases},\quad f_{Y|X}(y|x)=\begin{cases}2y/(1-x^4) & x^2\le y\le 1\\ 0 & \text{其他}\end{cases}$$

二维正态向量的条件分布

定理：若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则

$$X|Y=y\sim N(\mu_1+\rho(y-\mu_2)\sigma_1/\sigma_2, \sigma_1^2(1-\rho^2)),$$

$$Y|X=x\sim N(\mu_2+\rho(x-\mu_1)\sigma_2/\sigma_1, \sigma_2^2(1-\rho^2)).$$

证明：只须证明第一部分，第二部分由对称性可得。

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\sqrt{2\pi}\sigma_2}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times\\
&\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)+\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}\\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x-(\mu_1+\rho(y-\mu_2)\sigma_1/\sigma_2)\right]^2\right\}.
\end{aligned}
$$

目录

离散型随机向量和函数的分布

设二维离散型随机向量(X,Y)的分布列为

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

则$Z=X+Y$的分布列为

$$P(Z=z_k)=\sum_{x_i+y_j=z_k}p_{ij},\ k=1,2,\cdots.$$

特别地，如果(X,Y)的分布列为

$$P(X=i,Y=j)=p_{ij},\ i,j=0,1,2,\cdots.$$

则Z的分布列为

$$P(Z=k)=\sum_{i=1}^{k}P(X=i,Y=k-i)=\sum_{i=0}^{k}p_{i(k-i)},\ k=0,\cdots.$$

泊松分布的可加性

定理：如果$X\sim Pois(\lambda_1)$, $Y\sim Pois(\lambda_2)$, 且X与Y独立，则

$$X+Y\sim Pois(\lambda_1+\lambda_2).$$

证明：令$Z=X+Y$. 对任意非负整数k,

$$
\begin{aligned}
P(Z=k) &= \sum_{i=0}^{k}P(X=i,Y=k-i)=\sum_{i=0}^{k}P(X=i)P(Y=k-i)\\
&= \sum_{i=0}^{k}\frac{e^{-\lambda_1}\lambda_1^i}{i!}\frac{e^{-\lambda_2}\lambda_2^{k-i}}{(k-i)!} = e^{-\lambda_1-\lambda_2}\lambda_2^k\sum_{i=0}^{k}\frac{(\lambda_1/\lambda_2)^i}{i!(k-i)!}\\
&= \frac{e^{-\lambda_1-\lambda_2}\lambda_2^k}{k!}\sum_{i=0}^{k}\frac{(\lambda_1/\lambda_2)^i k!}{i!(k-i)!} = \frac{e^{-\lambda_1-\lambda_2}\lambda_2^k}{k!}(\lambda_1/\lambda_2+1)^k\\
&= \frac{(\lambda_1+\lambda_2)^k}{k!}e^{-\lambda_1-\lambda_2}
\end{aligned}
$$

二项分布的可加性

定理：如果$X \sim B(n,p)$，$Y \sim B(m,p)$，且X与Y独立，则

$$X + Y \sim B(n+m, p).$$

证明：令$Z = X + Y$. 对任意非负整数$k \le n+m$,

$$\begin{aligned} P(Z=k) &= \sum_{i=0}^{k} P(X=i)P(Y=k-i) \\ &= \sum_{i=0}^{k} C_n^i p^i (1-p)^{n-i} C_m^{k-i} p^{k-i} (1-p)^{m-k+i} \\ &= \left(\sum_{i=0}^{k} C_n^i C_m^{k-i} \right) p^k (1-p)^{n+m-k} \\ &= C_{n+m}^k p^k (1-p)^{n+m-k} \end{aligned}$$

连续型随机向量和函数的分布

设(X,Y)的联合密度为$f(x,y)$, 则$Z = X + Y$的分布函数为

$$\begin{aligned} F_Z(z) = P(Z \le z) &= \int_{x+y \le z} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{z-x} f(x,y)\,\mathrm{d}y \\ &= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f(x, y-x)\,\mathrm{d}x \right) \mathrm{d}y. \end{aligned}$$

求导得到Z的密度函数为

$$f_Z(z) = F_Z'(z) = \int_{-\infty}^{\infty} f(x, z-x)\,\mathrm{d}x.$$

特别地，如果X与Y独立，则

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_X(z-x) f_Y(x)\,\mathrm{d}x.$$

这就是卷积公式。

正态分布的可加性

定理：如果$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$，且$X$与$Y$独立，则

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

证明：令$X' = (X - \mu_1)/\sigma_2 \sim N(0, \tilde{\sigma}_1^2)$，其中$\tilde{\sigma}_1 = \sigma_1/\sigma_2$
$Y' = (Y - \mu_2)/\sigma_2 \sim N(0,1)$，$Z = X' + Y'$.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X'}(x) f_{Y'}(z-x)\,\mathrm{d}x \\ &= \frac{1}{2\pi\tilde{\sigma}_1} \int_{-\infty}^{\infty} \exp\left\{ -\frac{x^2}{2\tilde{\sigma}_1^2} - \frac{(z-x)^2}{2} \right\} \mathrm{d}x \\ &= \frac{1}{2\pi\tilde{\sigma}_1} \exp\left\{ -\frac{z^2}{2(\tilde{\sigma}_1^2+1)} \right\} \int_{-\infty}^{\infty} \exp\left\{ -\frac{(1+\tilde{\sigma}_1^2)(x - \tilde{\sigma}_1^2 z/(1+\tilde{\sigma}_1^2))^2}{2\tilde{\sigma}_1^2} \right\} \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\tilde{\sigma}_1^2+1}} \exp\left\{ -\frac{z^2}{2(\tilde{\sigma}_1^2+1)} \right\}. \end{aligned}$$

所以$Z \sim N(0, \tilde{\sigma}_1^2 + 1)$. $X + Y = \sigma_2 Z + \mu_1 + \mu_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

多维正态分布的线性组合

定理1：随机向量$\mathbf{X} = (X_1, \ldots, X_n)^\top$服从$n$维正态分布的充要条件是对任意的$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$，$\boldsymbol{\alpha}^\top \mathbf{X}$服从一维正态分布。

定理2：如果随机向量$\mathbf{X} = (X_1, \ldots, X_n)^\top \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则有对任意实矩阵$\mathbf{A} \in \mathbb{R}^{m \times n}$，$\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$.

注：如果X, Y都服从正态分布，(X,Y)不一定服从二维正态分布。（反例见作业题）

变量变换法

设二维随机向量(X,Y)的联合密度函数为$f_{X,Y}(x,y)$，如果函数

$$\begin{cases} u &= g_1(x,y) \\ v &= g_2(x,y) \end{cases}$$

有连续的偏导数，且存在唯一的反函数

$$\begin{cases} x &= x(u,v) \\ y &= y(u,v) \end{cases},$$

其变换的雅克比行列式

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}^{-1} \ne 0,$$

变量变换法

若

$$\begin{cases} U &= g_1(X,Y) \\ V &= g_2(X,Y) \end{cases},$$

则(U,V)的联合密度为

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v))\,|J|.$$

证：参见二重积分的变量变换法。

增补变量法

目的： 求$U = g(X, Y)$的密度函数

步骤1： 增补一个新的随机变量$V = h(X, Y)$ (一般令$V = X$或者$V = Y$)
步骤2： 用变量变换法求出(U, V)的联合密度函数

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))\,|J|$$

步骤3： 最后，通过联合密度函数$f_{U,V}(u, v)$求U的边缘分布

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u, v)\,\mathrm{d}v$$

注： 利用这种方法很容易求出的两个连续随机变量“积的分布”和“商的分布”

二维连续型随机向量积的分布

例； 已知(X, Y)的联合密度为$f(x, y)$，求$U = XY$的密度函数。

解： 设$V = X$. 则(U, V)的联合密度函数为

$$f_{U,V}(u, v) = f(v, u/v)\,|J|\,,$$

其中

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix} = -\frac{1}{v}$$

所以U的密度函数为

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u, v)\,\mathrm{d}v = \int_{-\infty}^{\infty} f(v, u/v)\frac{1}{|v|}\,\mathrm{d}v.$$

二维连续型随机向量商的分布

例； 已知(X, Y)的联合密度为$f(x, y)$，求$U = X/Y$的密度函数。

解： 设$V = Y$. 则(U, V)的联合密度函数为

$$f_{U,V}(u, v) = f(uv, v)\,|J|\,,$$

其中

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

所以U的密度函数为

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u, v)\,\mathrm{d}v = \int_{-\infty}^{\infty} f(uv, v)\,|v|\,\mathrm{d}v.$$

第三次作业

设(X, Y)的联合密度函数为

$$f(x, y) = \frac{3}{8\pi\sqrt{2}}\left\{\exp\left(-\frac{9}{16}(x^2 - \frac{2}{3}xy + y^2)\right) + \exp\left(-\frac{9}{16}(x^2 + \frac{2}{3}xy + y^2)\right)\right\}$$

- 求X, Y的边缘分布并判断(X, Y)是否服从二维正态分布
- 分别求$X + Y$, X/Y的密度函数

第三章习题：4, 8, 10, 14, 15, 19

5月9日（周三）交！

概率论与数理统计

何志坚

华南理工大学数学学院
hezhijian@scut.edu.cn

第四章：随机变量的数字特征

随机变量的自我介绍

随机变量X：本人随机变量X，连续型，期望为0.1，方差为1.2，偏度为-1.1，峰度为2.2....

随机变量Y：我是随机变量Y，也是连续型，与随机变量X是好朋友，我们两的相关系数是0.99. 我觉得X是个很好的随机变量！

Dr. He: 好与坏关键要看是否“合适”。

目录

一维随机变量的数字特征

概率分布全面地描述了随机变量的统计规律性，由分布可以计算有关随机变量事件的概率。除此之外，由分布还可以计算随机变量的一些数字特征。这些特征从侧面描述了分布的特征。

- 期望：描述分布的“平均水平”
- 方差：描述分布的“波动大小”

环数	8	9	10
甲	0.1	0.8	0.1
乙	0.8	0.1	0.1
丙	0.2	0.6	0.2

如何比较甲、乙、丙三个选手的水平高低?

离散型随机变量数学期望的定义

定义：设离散随机变量X的分布列为

$$p_i = p(x_i) = P(X = x_i),\ i = 1, 2, \ldots.$$

如果级数$\sum_{i=1}^{\infty} |x_i|\, p(x_i) < \infty$[1], 则称$X$数学期望存在，并称

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

为X的数学期望(Expectation)，简称X的期望。

环数	8	9	10	期望
甲	0.1	0.8	0.1	$8 \times 0.1 + 9 \times 0.8 + 10 \times 0.1 = 9.0$
乙	0.8	0.1	0.1	$8 \times 0.8 + 9 \times 0.1 + 10 \times 0.1 = 8.3$
丙	0.2	0.6	0.2	$8 \times 0.2 + 9 \times 0.6 + 10 \times 0.2 = 9.0$

[1] 要求级数绝对收敛的目的在于使数学期望唯一。

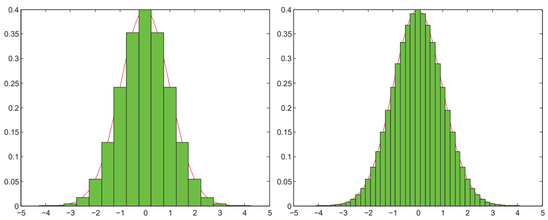
连续型随机变量数学期望的定义

定义：设连续型随机变量X的密度函数为$p(x)$. 如果

$$\int_{-\infty}^{\infty} |x|\, p(x)\, \mathrm{d}x < \infty,$$

则称X数学期望存在，定义为

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x p(x)\, \mathrm{d}x.$$



随机变量函数的数学期望

定理：随机变量X的某一函数$g(X)$的数学期望为

$$\mathbb{E}[g(X)] = \begin{cases} \sum_{i=1}^{\infty} g(x_i) p(x_i) & \text{离散型} \\ \int_{-\infty}^{\infty} g(x) p(x)\, \mathrm{d}x & \text{连续型} \end{cases}$$

期望的性质：

- 如果c为常数，则$\mathbb{E}(c) = c$.
- 对任意常数a, 有$\mathbb{E}(aX) = a\mathbb{E}(X)$.
- 对于任意两个函数g_1, g_2, 有

$$\mathbb{E}[g_1(X) \pm g_2(X)] = \mathbb{E}[g_1(X)] \pm \mathbb{E}[g_2(X)].$$

- 如果$a \le X \le b$, 则$\mathbb{E}(X)$存在，且$a \le \mathbb{E}(X) \le b$.

随机变量的方差与标准差

定义：设X为一随机变量。如果$\mathbb{E}(X^2)<\infty$, 则称偏差平方$(X-\mathbb{E}(X))^2$的数学期望$\mathbb{E}[(X-\mathbb{E}(X))^2]$为随机变量$X$的方差(Variance)，记为

$$\mathrm{Var}(X)=D(X)=\mathbb{E}[(X-\mathbb{E}(X))^2]=\begin{cases}\sum_{i=1}(x_i-\mathbb{E}(X))^2p(x_i) & \text{离散型}\\ \int_{-\infty}^{\infty}(x-\mathbb{E}(X))^2p(x)\,\mathrm{d}x & \text{连续型}.\end{cases}$$

称方差的正平方根$\sqrt{\mathrm{Var}(X)}$为标准差，记为$\sigma(X)$或σ_X.

注：方差与标准差用来描述随机变量取值的集中与分散程度，如果他们越小则说明随机变量的取值越集中；反之越发散。

环数	8	9	10	期望	方差
甲	0.1	0.8	0.1	9.0	$1^2\times0.1+1^2\times0.1=0.20$
乙	0.8	0.1	0.1	8.3	$0.3^2\times0.8+0.7^2\times0.1+1.7^2\times0.1=0.41$
丙	0.2	0.6	0.2	9.0	$1^2\times0.2+1^2\times0.2=0.40$

方差的性质

方差的等价表达形式

$$\mathrm{Var}(X)=\mathbb{E}(X^2)-[\mathbb{E}(X)]^2$$

- 如果c为常数，则$\mathrm{Var}(c)=0$, $\mathrm{Var}(X+c)=\mathrm{Var}(X)$.
- 对任意常数a, 有$\mathrm{Var}(aX)=a^2\mathrm{Var}(X)$.
- $\mathrm{Var}(X)=0$的等价条件是存在常数c使得$P(X=c)=1$.

二项分布的期望与方差

定理：如果$X\sim B(n,p)$, 则$\mathbb{E}[X]=np$, $\mathrm{Var}[X]=np(1-p)$.

证：令$q=1-p$.

$$\begin{aligned}\mathbb{E}[X]&=\sum_{k=0}^{n}kB(k;n,p)=\sum_{k=1}^{n}kC_n^kp^kq^{n-k}\\&=np\sum_{k=1}^{n}C_{n-1}^{k-1}p^{k-1}q^{n-k}\\&=np\sum_{k=0}^{n-1}C_{n-1}^{k}p^kq^{n-1-k}=np\end{aligned}$$

其中用到$kC_n^k=nC_{n-1}^{k-1}$

二项分布的期望与方差

定理：如果$X\sim B(n,p)$, 则$\mathbb{E}[X]=np$, $\mathrm{Var}[X]=np(1-p)$.

证：

$$\begin{aligned}\mathbb{E}[X^2]&=np\sum_{k=1}^{n}kC_{n-1}^{k-1}p^{k-1}q^{n-k}\\&=np\sum_{k=0}^{n-1}(k+1)C_{n-1}^{k}p^kq^{n-1-k}\\&=np\left(\sum_{k=1}^{n-1}kC_{n-1}^{k}p^kq^{n-1-k}+1\right)\\&=np((n-1)p+1)=(np)^2+npq.\end{aligned}$$

其中用到$\sum_{i=1}^n iC_n^ip^{i-1}q^{n-i}=n(p+q)^{n-1}$. 所以

$$\mathrm{Var}[X]=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=npq=np(1-p).$$

泊松分布的期望与方差

定理：如果$X\sim Pois(\lambda)$, 则$\mathbb{E}[X]=\mathrm{Var}[X]=\lambda$.

证：

$$\begin{aligned}\mathbb{E}[X]&=e^{-\lambda}\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=1}^{\infty}k\frac{\lambda^k}{k!}\\&=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\lambda.\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X^2]&=e^{-\lambda}\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}=\lambda e^{-\lambda}\sum_{k=0}^{\infty}(k+1)\frac{\lambda^k}{k!}\\&=\lambda e^{-\lambda}\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}+\lambda=\lambda^2+\lambda\end{aligned}$$

$$\mathrm{Var}[x]=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=\lambda$$

几何分布的期望与方差

定理：如果$X\sim Geo(p)$, 则$\mathbb{E}[X]=\frac{1}{p}$, $\mathrm{Var}[X]=\frac{1-p}{p^2}$.

证：设$q=1-p$.

$$\mathbb{E}[X]=\sum_{k=1}^{\infty}kq^{k-1}p=p\frac{1}{(1-q)^2}=\frac{1}{p}.$$

$$\begin{aligned}\mathbb{E}[X^2]&=p\sum_{k=1}^{\infty}k^2q^{k-1}=p\sum_{k=1}^{\infty}[k(k-1)+k]q^{k-1}\\&=pq\sum_{k=2}^{\infty}k(k-1)q^{k-2}+\frac{1}{p}\\&=pq\frac{2}{(1-q)^3}+\frac{1}{p}=\frac{2q}{p^2}+\frac{1}{p}\end{aligned}$$

$$\mathrm{Var}[x]=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=\frac{2q}{p^2}+\frac{1}{p}-\frac{1}{p^2}=\frac{q}{p^2}$$

均匀分布的期望与方差

定理：如果$X \sim \mathbb{U}[a,b]$, 则$\mathbb{E}[X]=(b-a)/2$, $\mathrm{Var}[X]=(b-a)^2/12$.

证：

$$\mathbb{E}[X]=\int_a^b \frac{x}{b-a}\,\mathrm{d}x=\frac{a+b}{2}$$

$$\mathbb{E}[X^2]=\int_a^b \frac{x^2}{b-a}\,\mathrm{d}x=\frac{a^2+ab+b^2}{3}$$

$$\mathrm{Var}[x]=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=\frac{a^2+ab+b^2}{3}-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}$$

指数分布的期望与方差

定理：如果$X \sim Exp(\lambda)$, 则$\mathbb{E}[X]=1/\lambda$, $\mathrm{Var}[X]=(b-a)^2/12$.

证：

$$\begin{aligned}\mathbb{E}[X]&=\int_0^\infty x\lambda e^{-\lambda x}\,\mathrm{d}x=-\int_0^\infty x\,\mathrm{d}e^{-\lambda x}\\&=-xe^{-\lambda x}|_0^\infty+\int_0^\infty e^{-\lambda x}\,\mathrm{d}x=\frac{1}{\lambda}\ \text{(分部积分)}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X^2]&=\int_0^\infty x^2\lambda e^{-\lambda x}\,\mathrm{d}x\\&=-x^2e^{-\lambda x}|_0^\infty+2\int_0^\infty xe^{-\lambda x}\,\mathrm{d}x\ \text{(分部积分)}\\&=\frac{2}{\lambda^2}\end{aligned}$$

$$\mathrm{Var}[x]=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=\frac{1}{\lambda^2}$$

正态分布的期望与方差

定理：如果$X \sim N(\mu,\sigma^2)$, 则$\mathbb{E}[X]=\mu$, $\mathrm{Var}[X]=\sigma^2$.

证：

$$\mathbb{E}[X-\mu]=\int_{-\infty}^{\infty}(x-\mu)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x=0.$$

所以$\mathbb{E}[X]=\mu$.

$$\begin{aligned}\mathrm{Var}[x]&=\int_{-\infty}^{\infty}(x-\mu)^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x\\&=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2e^{-x^2/2}\,\mathrm{d}x\\&=-\frac{\sigma^2}{\sqrt{2\pi}}xe^{-x^2/2}|_{-\infty}^{\infty}+\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}\,\mathrm{d}x\ \text{(分部积分)}\\&=\sigma^2.\end{aligned}$$

柯西分布的期望与方差

定理：柯西分布的期望与方差不存在。

证： 设X服从柯西分布，其密度函数为

$$f(x)=\frac{1}{\pi(1+x^2)},\ x\in\mathbb{R}.$$

$$\begin{aligned}\mathbb{E}[|X|]&=\int_{-\infty}^{\infty}\frac{|x|}{\pi(1+x^2)}\,\mathrm{d}x=\frac{2}{\pi}\int_0^\infty\frac{x}{1+x^2}\,\mathrm{d}x\\&=\frac{1}{\pi}\int_0^\infty\frac{1}{1+y}\,\mathrm{d}y\\&=\frac{1}{\pi}\int_1^\infty\frac{1}{y}\,\mathrm{d}y=\infty\end{aligned}$$

总结

分布名称	记号	期望	方差
二项分布	$B(n,p)$	np	npq
泊松分布	$Pois(\lambda)$	λ	λ
几何分布	$Geo(p)$	$1/p$	q/p^2
均匀分布	$U[a,b]$	$(a+b)/2$	$(b-a)^2/12$
指数分布	$Exp(\lambda)$	$1/\lambda$	$1/\lambda^2$
正态分布	$N(\mu,\sigma^2)$	μ	σ^2
柯西分布		不存在	不存在

作业：计算上述前六种分布的期望与方差。5月11日(周五)交！

随机变量的矩

定义：设X为随机变量，c为常数，k为正整数，如果$\mathbb{E}[|X-c|^k]<\infty$, 则称

$$\mathbb{E}[(X-c)^k]$$

为X关于点c的k阶矩。特别地，

(1) 当$c=0$，称$\mathbb{E}[X^k]$为X的k阶原点矩；

(2) 当$c=\mathbb{E}[X]$，称$\mathbb{E}[(X-\mathbb{E}[X])^k]$为$X$的$k$阶中心矩。

注：期望是一阶原点矩；方差是二阶中心矩。

标准正态分布的矩

定理：如果$X \sim N(0,1)$，对任意的正整数k，有

$$\mathbb{E}[X^k] = \begin{cases} 0 & k\text{为奇数} \\ (k-1)!! & k\text{为偶数} \end{cases}$$

证：当k为奇数时，$\mathbb{E}[X^k] = \mathbb{E}[(-X)^k] = -\mathbb{E}[X^k]$. 所以$\mathbb{E}[X^k] = 0$. 当$k$为偶数时，

$$\begin{aligned}\mathbb{E}[X^k] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2}\,\mathrm{d}x = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k-1}\,\mathrm{d}e^{-x^2/2} \\ &= -\frac{1}{\sqrt{2\pi}} x^{k-1} e^{-x^2/2}\Big|_{-\infty}^{\infty} + \frac{k-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k-2} e^{-x^2/2}\,\mathrm{d}x \\ &= (k-1)\mathbb{E}[X^{k-2}]\end{aligned}$$

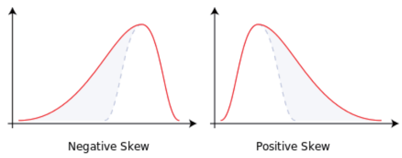
所以$\mathbb{E}[X^k] = (k-1)!!\mathbb{E}[X^0] = (k-1)!!$.

随机变量的偏度

定义：设X为随机变量，如果$\mathbb{E}[|X|^3] < \infty$且$\mathrm{Var}[x] > 0$, 则称

$$\beta_s := \frac{\mathbb{E}[(X-\mathbb{E}[X])^3]}{(\mathrm{Var}[X])^{3/2}} = \mathbb{E}\left[\left(\frac{X-\mathbb{E}[X]}{\sigma_X}\right)^3\right]$$

为X的偏度(skewness). 如果$\beta_s < 0$时，称该分布左(负)偏；如果$\beta_s > 0$时，称该分布右(正)偏。



注：偏度β_s是刻画分布偏离对称性程度的一个特征数。定义中的分母的作用是为了消除量纲的影响。

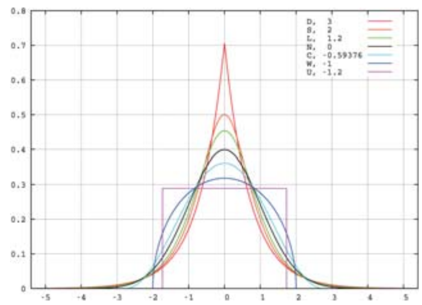
随机变量的峰度

定义：设X为随机变量，如果$\mathbb{E}[X^4] < \infty$且$\mathrm{Var}[x] > 0$, 则称下式为X的峰度(kurtosis)：

$$\beta_k = \frac{\mathbb{E}[(X-\mathbb{E}[X])^4]}{(\mathrm{Var}[X])^2} = \mathbb{E}\left[\left(\frac{X-\mathbb{E}[X]}{\sigma_X}\right)^4\right].$$

注：偏度β_k是刻画分布尖峭程度和（或）尾部粗细的一个特征数。通常与标准正态的峰度3比较。如果$\beta_k > 3$, 说明标准化后的分布比标准正态分布更尖峭和（或）尾部更粗；如果$\beta_k < 3$, 说明标准化后的分布比标准正态分布更平坦和（或）尾部更细。如果$\beta_s \approx 0, \beta_k \approx 3$, 常认为该分布为近似正态分布。

不同分布的峰度



图：图中为7种期望为0且方差为1的不同分布示意图，其他U表示均匀分布，N表示标准正态分布。右上角的数值表示$\beta_k - 3$的值

分位数

定义：设随机变量X的分布函数为$F(x)$, 密度函数为$p(x)$. 对任意$p \in (0,1)$, 称满足条件

$$F(x_p) = \int_{-\infty}^{x_p} p(x)\,\mathrm{d}x = p$$

的x_p为此分布的p分位数(quantile). 当$p = 0.5$时，x_p称为中位数。

例：标准正态分布的分位数：

p	0.01	0.05	0.1	0.5	0.9	0.95	0.99
x_p	-2.33	-1.64	-1.28	0	1.28	1.64	2.33

注意到$x_{1-p} = -x_p$

目录

二维随机向量函数的方差与期望

推广：二维随机向量(X,Y)的某一函数$g(X,Y)$的数学期望为

$$\mathbb{E}[g(X,Y)] = \begin{cases} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)P(X=x_i,Y=y_j) & \text{离散型} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p(x,y)\,\mathrm{d}x\,\mathrm{d}y & \text{连续型} \end{cases}$$

类似地，

$$\mathrm{Var}[g(X,Y)] = \mathbb{E}[(g(X,Y)-\mathbb{E}[g(X,Y)])^2] = \mathbb{E}[g(X,Y)^2] - (\mathbb{E}[g(X,Y)])^2$$

性质：

- 线性性：$\mathbb{E}[aX+bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
- 如果X,Y独立，$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, $\mathrm{Var}[aX+bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y]$.
- 上述定义以及性质可以推广到n维随机向量。

二维随机向量的协方差

定义：设(X,Y)为二维随机向量，且$\mathrm{Var}[X]>0$, $\mathrm{Var}[Y]>0$. 其协方差(covariance)定义为

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])].$$

相关系数定义为

$$r(X,Y) = \rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}.$$

如果$r(X,Y)=0$, 称X,Y不相关。

协方差的性质(I)

- 等价表达形式：
$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$
- 对称性：
$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$$
- 双线性性：
$$\mathrm{Cov}(aX+bY,Z) = a\mathrm{Cov}(X,Z) + b\mathrm{Cov}(Y,Z)$$
$$\mathrm{Cov}(aX,bY) = ab\mathrm{Cov}(X,Y)$$
-
$$\mathrm{Var}[X] = \mathrm{Cov}(X,X)$$
$$\mathrm{Var}(X+Y) = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y)$$
- 如果X,Y独立，$\mathrm{Var}(X,Y)=0$.

施瓦茨(Schwarz)不等式

定理：设(X,Y)为二维随机向量，则有

$$(\mathbb{E}[XY])^2 \le \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

证：令

$$g(t) = \mathbb{E}[(tX+Y)^2] = \mathbb{E}[X^2]t^2 + 2\mathbb{E}[XY]t + \mathbb{E}[Y^2] \ge 0.$$

如果$\mathbb{E}[X^2]=0$, 则意味着$P(X=0)=1$. 所以$\mathbb{E}[XY]=0$.
如果$\mathbb{E}[X^2]>0$, 则$g(t)$的判别式不大于0，即

$$(2\mathbb{E}[XY])^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2] \le 0.$$

推论：设(X,Y)为二维随机向量，则有

$$[\mathrm{Cov}(X,Y)]^2 \le \mathrm{Var}[X]\mathrm{Var}[Y].$$

相关系数的性质

定理：设(X,Y)为二维随机向量，且$\mathrm{Var}[X]>0$, $\mathrm{Var}[Y]>0$.

- 如果X,Y独立，则$r(X,Y)=0$, 即X,Y不相关。
-
$$|r(X,Y)| \le 1$$
- $r(X,Y)=1$的充要条件是存在常数$a>0$和b, 使得
$$P(Y=aX+b)=1$$
$r(X,Y)=-1$的充要条件是存在常数$a<0$和b, 使得
$$P(Y=aX+b)=1$$

证：注意到$\mathrm{Var}(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}) = 2[1 \pm r(X,Y)]$
注：相关系数反映X与Y线性相关的程度。若$r=1$，称X,Y正相关，若$r=-1$，称X,Y负相关。

二维正态分布的相关系数

例：设$(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. 则

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

证明：

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mu_1)(Y-\mu_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xye^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}\,\mathrm{d}x\,\mathrm{d}y = \rho\sigma_1\sigma_2.$$

$$r(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_1\sigma_2} = \rho$$

注：由此可得，对于二维正态分布，独立性与不相关性是等价的。但一般情况下，这种等价关系不存在（独立可以得到不相关，反之不成立）

反例

设(X,Y)的联合密度函数为

$$f(x,y)=\frac{3}{8\pi\sqrt{2}}\left\{\exp\left(-\frac{9}{16}(x^2-\frac{2}{3}xy+y^2)\right)+\exp\left(-\frac{9}{16}(x^2+\frac{2}{3}xy+y^2)\right)\right\}$$

求$r(X,Y)$.

解：易知，$X\sim N(0,1), Y\sim N(0,1)$. 容易发现$\mathbb{E}[XY]=0$.
又$\mathbb{E}[X]=\mathbb{E}[Y]=0$. 由此可得$r(X,Y)=0$.

注：虽然$r(X,Y)=0$，但X,Y不独立。这是因为$f(x,y)\neq f_X(x)f_Y(y)$.

例

例：已知X,Y相互独立，均服从正态分布$N(0,\sigma^2)$，$U=aX+bY$，$V=aX-bY$，其中a,b是常数。

- 求U,V的相关系数
- U,V是否相关，是否独立?
- 当U,V独立时，求它们的联合密度函数。

解：$\mathbb{E}[UV]=\mathbb{E}[(aX+bY)(aX-bY)]=a^2\mathbb{E}[X^2]-b^2\mathbb{E}[Y^2]=(a^2-b^2)\sigma^2$. 由于$\mathbb{E}[U]=\mathbb{E}[V]=0$, 所以$\mathrm{Cov}(X,Y)=(a^2-b^2)\sigma^2$.

$$\mathrm{Var}(U)=\mathrm{Var}(V)=a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)=(a^2+b^2)\sigma^2.$$

$$r(X,Y)=\frac{a^2-b^2}{a^2+b^2}$$

协方差矩阵

定义：设随机向量$\mathbf{X}=(X_1,\ldots,X_n)^\top$的每个分量都有有限方差，定义协方差矩阵如下

$$\mathrm{Var}(\mathbf{X})=\begin{pmatrix}\mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & \cdots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & \vdots & \vdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Cov}(X_n,X_n)\end{pmatrix}$$

相关系数矩阵定义如下：

$$r(\mathbf{X})=\begin{pmatrix}1 & r(X_1,X_2) & \cdots & r(X_1,X_n)\\ r(X_2,X_1) & 1 & \cdots & r(X_2,X_n)\\ \vdots & \vdots & \vdots & \vdots\\ r(X_n,X_1) & r(X_n,X_2) & \cdots & 1\end{pmatrix}$$

条件数学期望

定义：设(X,Y)为二维随机向量，有有限的数学期望。在$\{Y=y\}$发生的条件下，X的条件数学期望（简称条件期望），就是在条件分布$X|Y=y$下求条件期望。

$$\mathbb{E}[X|Y=y]=\begin{cases}\sum_{i=1}^{\infty}x_iP(X=x_i|Y=y) & \text{离散型}\\ \int_{-\infty}^{\infty}xf_{X|Y}(x|y)\,\mathrm{d}x & \text{连续型}\end{cases}$$

同样方式定义$\mathbb{E}[Y|X=x]$.

注1：$\mathbb{E}[X|Y=y]$可以看成关于y的函数，记为$g(y)=\mathbb{E}[X|Y=y]$.
则$g(Y)=\mathbb{E}[X|Y]$为随机变量Y的函数。同样$\mathbb{E}[Y|X]$为随机变量X的函数。
注2：如果X,Y独立$\mathbb{E}[X|Y]=\mathbb{E}[X]$，$\mathbb{E}[Y|X]=\mathbb{E}[Y]$.
注3：条件期望满足前面提到期望的所有性质，比如线性性。

全期望公式

$$\mathbb{E}[\mathbb{E}[X|Y]]=\mathbb{E}[X]$$

证明：考虑离散的情形。

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]]&=\sum_{j=1}^{\infty}\mathbb{E}[X|Y=y_j]P(Y=y_j)\\&=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_iP(X=x_i|Y=y_j)P(Y=y_j)\\&=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_iP(X=x_i,Y=y_j)\\&=\sum_{i=1}^{\infty}x_i\sum_{j=1}^{\infty}P(X=x_i,Y=y_j)\\&=\sum_{i=1}^{\infty}x_iP(X=x_i)=\mathbb{E}[X]\end{aligned}$$

全期望公式

$$\mathbb{E}[\mathbb{E}[X|Y]]=\mathbb{E}[X]$$

证明：考虑连续的情形。

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]]&=\int_{-\infty}^{\infty}\mathbb{E}[X|Y=y]f_Y(y)\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{X|Y}(x|y)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}x\left(\int_{-\infty}^{\infty}f_{X,Y}(x,y)\,\mathrm{d}y\right)\mathrm{d}x\\&=\int_{-\infty}^{\infty}xf_X(x)\,\mathrm{d}x=\mathbb{E}[X]\end{aligned}$$

例1：二维正态向量的条件期望

例：若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求$\mathbb{E}[X|Y],\mathbb{E}[Y|X]$.

解：前一章已经证明：

$$X|Y=y\sim N(\mu_1+\rho(y-\mu_2)\sigma_1/\sigma_2,\sigma_1^2(1-\rho^2)),$$

$$Y|X=x\sim N(\mu_2+\rho(x-\mu_1)\sigma_2/\sigma_1,\sigma_2^2(1-\rho^2)).$$

所以，$\mathbb{E}[X|Y]=\mu_1+\rho(Y-\mu_2)\sigma_1/\sigma_2$. $\mathbb{E}[Y|X]=\mu_2+\rho(X-\mu_1)\sigma_2/\sigma_1$.

例2：随机个随机变量的和

例：设$X_1,X_2,\cdots$为独立同分布随机变量序列，期望为μ_1，N为非负整数值随机变量，期望为μ_2，且与$X_1,X_2,\cdots$独立，求$\mathbb{E}[\sum_{k=1}^N X_k]$.

解：由全期望公式可得，

$$\begin{aligned}\mathbb{E}\left[\sum_{k=1}^N X_k\right]&=\mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^N X_k\middle|N\right]\right]\\&=\sum_{n=0}^\infty\mathbb{E}\left[\sum_{k=1}^N X_k\middle|N=n\right]P(N=n)\\&=\sum_{n=0}^\infty\mathbb{E}\left[\sum_{k=1}^n X_k\right]P(N=n)\\&=\sum_{n=0}^\infty\mu_1 nP(N=n)=\mu_1\sum_{n=0}^\infty nP(N=n)=\mu_1\mu_2.\end{aligned}$$

条件方差

定义：条件方差定义：

$$\mathrm{Var}[X|Y]=\mathbb{E}[(\mathbb{E}[X|Y]-\mathbb{E}[X])^2]=\mathbb{E}[(\mathbb{E}[X|Y])^2]-(\mathbb{E}[X])^2$$

性质：

$$\mathrm{Var}[X]=\mathrm{Var}[\mathbb{E}[X|Y]]+\mathbb{E}[\mathrm{Var}[X|Y]]$$

作业

第四章习题：2, 8, 13, 14, 16, 21, 22

5月18日（周五）交！

概率论与数理统计

何志坚

华南理工大学数学学院

hezhijian@scut.edu.cn

第五章：大数定律和中心极限定理

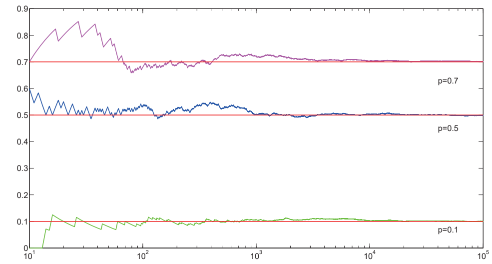
目录

试验1

考虑n重伯努利试验，事件A发生的概率为p. 记

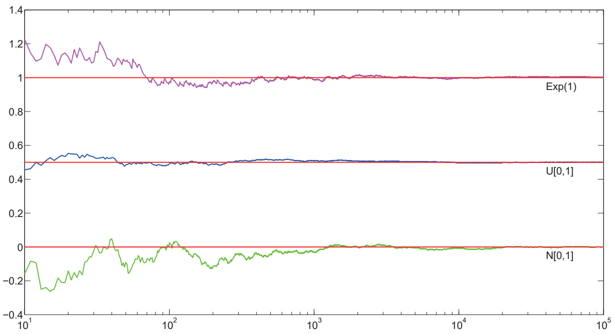
$$X_i = \begin{cases} 1, & \text{第}i\text{次试验事件}A\text{发生} \\ 0, & \text{第}i\text{次试验事件}A\text{不发生} \end{cases}$$

令$S_n = X_1 + X_2 + \cdots + X_n$. 试问$S_n/n$是否会收敛到一个常数?



试验2

对于一般的随机序列X_i呢? 试问S_n/n是否会收敛到一个常数?



随机变量序列的收敛

定义：如果对任意$\epsilon > 0$有

$$\lim_{n\to\infty} P(\omega : |Y_n(\omega) - Y(\omega)| \geq \epsilon) = 0,$$

则称Y_n依概率收敛于Y，记作$Y_n \xrightarrow{\mathrm{P}} Y$.

定义：如果对任意$\epsilon > 0$有

$$P\left(\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right) = 1,$$

则称Y_n以概率1收敛(又称几乎处处收敛)于Y，记作$Y_n \xrightarrow{\mathrm{a.s}} Y$.

注：概率为0不等于不可能，概率为1不等于一定。

形象解释：开始上课了，慢慢地大家都安静下来，这是几乎处处收敛。绝大多数同学都安静下来，但每一个人都在不同的时间捣乱，这是依概率收敛。

依概率收敛与以概率1收敛的联系

定理：如果$Y_n \xrightarrow{\mathrm{a.s}} Y$，则$Y_n \xrightarrow{\mathrm{P}} Y$.

证明：令

$$C = \{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}.$$

则$P(C) = 1$. 令$A_n = \{\omega : |Y_n(\omega) - Y(\omega)| \geq \epsilon\}$,

$$B_n = \bigcup_{i=n}^{\infty} A_n,$$

$B_\infty = \cap_{n=1}^{\infty} B_n$. 容易发现$C \cap B_\infty = \varnothing$. 所以，$B_\infty \subset \bar{C}$, $P(B_\infty) \leq P(\bar{C}) = 0$, 即$P(B_\infty) = 0$.

$$\lim_{n\to\infty} P(A_n) \leq \lim_{n\to\infty} P(B_n) = P(B_\infty) = 0.$$

注：弱大数定律涉及依概率收敛，强大数定律涉及以概率1收敛。

伯努利弱大数定律

定理：设S_n为n重伯努利试验中事件A发生的次数，p为事件A出现的概率。则对任意$\epsilon > 0$有

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) = 0.$$

即$S_n/n \xrightarrow{\mathrm{P}} p$.

注：历史上，伯努利是第一个研究弱大数定理的，他在1713年发表的论文中，提出了上述定理，那是概率论的第一篇论文。

切比雪夫弱大数定律

定理：设$\{X_n\}$为两两不相关随机变量序列。若每个X_i的方差存在且有共同的上界，即$\mathrm{Var}[X_i] \le C$, 则对任意$\epsilon > 0$有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n \mathbb{E}[X_i]}{n}\right| \ge \epsilon\right) = 0.$$

特别地，如果所有的期望相等，即$\mathbb{E}[X_i] = \mu$，则

$$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{\mathrm{P}} \mu.$$

注：伯努利弱大数定律是切比雪夫弱大数定律的特例，其中

$$X_i = \begin{cases} 1, & 第i次试验事件A发生 \\ 0, & 第i次试验事件A不发生 \end{cases}$$

$\mathrm{Var}[X_i] = p(1-p)$.

切比雪夫不等式

引理：设k为正整数，随机变量X的k阶原点矩存在，则对任意$\epsilon > 0$有

$$P(|X| \ge \epsilon) \le \frac{\mathbb{E}[|X|^k]}{\epsilon^k}.$$

切比雪夫不等式：如果随机变量X的方差存在，则对任意$\epsilon > 0$有

$$P(|X - \mathbb{E}[X]| \ge \epsilon) \le \frac{\mathrm{Var}[X]}{\epsilon^2}.$$

证明：在引理中，替换X为$X - \mathbb{E}[X]$并取$k = 2$可得到切比雪夫不等式。只需证引理即可。

$$\begin{aligned} P(|X| \ge \epsilon) &= \mathbb{E}[1_{\{|X|\ge\epsilon\}}] \\ &\le \mathbb{E}\left[\frac{|X|^k}{\epsilon^k} 1_{\{|X|\ge\epsilon\}}\right] \\ &\le \mathbb{E}\left[\frac{|X|^k}{\epsilon^k}\right] = \frac{\mathbb{E}[|X|^k]}{\epsilon^k}. \end{aligned}$$

切比雪夫弱大数定律的证明

由切比雪夫不等式知，只需证$\lim_{n\to\infty} \mathrm{Var}[(\sum_{i=1}^n X_n)/n] = 0$.

$$\begin{aligned} \mathrm{Var}\left[\frac{\sum_{i=1}^n X_n}{n}\right] &= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^n X_n\right] \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left[X_n\right] (\text{由于两两不相关}) \\ &\le \frac{nC}{n^2} = \frac{C}{n} \to 0. \end{aligned}$$

马尔科夫大数定律

定理：如果

$$\lim_{n\to\infty} \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^n X_i\right] = 0,$$

则

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n \mathbb{E}[X_i]}{n}\right| \ge \epsilon\right) = 0.$$

辛钦大数定律

定理：设$\{X_n\}$为独立同分布的随机变量序列，具有有限的期望μ，则对任意$\epsilon > 0$有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| \ge \epsilon\right) = 0.$$

注：辛钦大数定律不需要方差存在，但要求独立同分布。

博雷尔(Borel)强大数定律

定理：设S_n为n重伯努利试验中事件A发生的次数，p为事件A出现的概率。则

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = p\right) = 1.$$

即$S_n/n \xrightarrow{\mathrm{a.s}} p$.

注：博雷尔大数定律为伯努利大数定律的“加强”版本。

柯尔莫哥洛夫强大数定律

定理1：设$\{X_n\}$为独立的随机变量序列，具有有限的期望，且

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}[X_n]}{n^2}<\infty,$$

则

$$P\left(\lim_{n\to\infty}\frac{\sum_{i=1}^{n}(X_i-\mathbb{E}[X_i])}{n}=0\right)=1.$$

定理2：设$\{X_n\}$为独立同分布的随机变量序列，具有有限的期望μ，则

$$P\left(\lim_{n\to\infty}\frac{\sum_{i=1}^{n}X_i}{n}=\mu\right)=1.$$

注：由定理2可以得到辛钦大数定律。

例

例：设$\{X_n\}$为独立同分布的随机变量序列，且$\mathbb{E}[X_n]=2, \mathrm{Var}[X_n]=6$，证明当$n\to\infty$，

$$\frac{X_1^2+X_2X_3+X_4^2+X_5X_6+\cdots+X_{3n-2}^2+X_{3n-1}X_{3n}}{n}\xrightarrow{\mathrm{P}}a,$$

并确定常数a的值。

证:令$Y_k=X_{3k-2}^2+X_{3k-1}X_{3k}$，则$\{Y_k\}$为独立同分布的随机变量序列。又

$$\mathbb{E}[Y_k]=\mathbb{E}[X_{3k-2}^2]+\mathbb{E}[X_{3k-1}]\mathbb{E}[X_{3k}]=\mathrm{Var}[X_1^2]+2(\mathbb{E}[X_1])^2=14.$$

由辛钦大数定律得，$(Y_1+\cdots+Y_n)/n\xrightarrow{\mathrm{P}}14$. 所以$a=14$.

目录

中心极限定理

考虑随机变量序列$\{X_n\}$，大数定律研究

$$\bar{X}_n:=\frac{X_1+X_2+\cdots+X_n}{n}$$

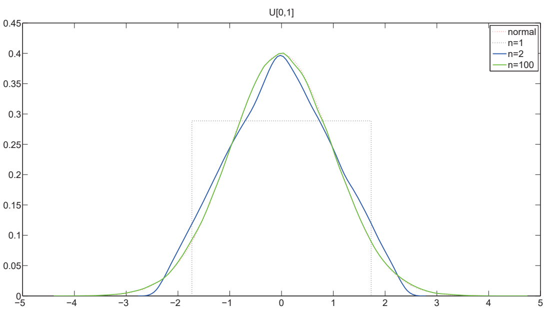
的收敛。中心极限定理研究$\bar{X}_n$的标准化后极限分布，即

$$\xi_n:=\frac{\bar{X}_n-\mathbb{E}[\bar{X}_n]}{\sigma(\bar{X}_n)}=\frac{\sum_{i=1}^{n}(X_i-\mathbb{E}[X_i])}{\sqrt{\mathrm{Var}[\sum_{i=1}^{n}X_i]}}$$

中心极限定理将告诉我们只要X_n独立同分布且方差有限，ξ_n的极限分布均为标准正态分布，与X_n的分布无关。

注：“统计学之父”卡尔·皮尔逊(Karl Pearson,1857–1936)教授认为：正态分布是上帝赐给人们唯一正确的分布。

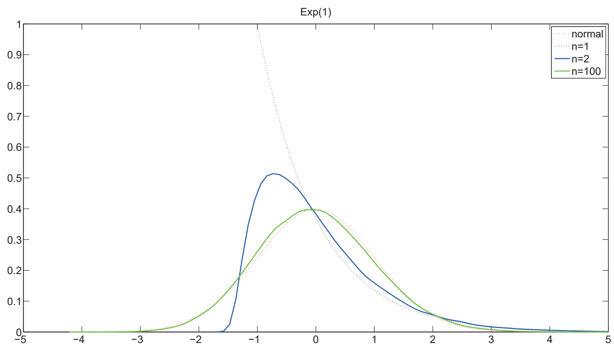
均匀分布



注：通过12个独立同分布的标准均匀分布可以近似得到标准正态分布，即

$$Y=\sum_{i=1}^{12}X_i-6\approx N(0,1)$$

指数分布



依分布收敛

定义：设Y是连续型随机变量。如果对任意$x \in \mathbb{R}$[1]都有

$$\lim_{n\to\infty} F_{Y_n}(x) = F_Y(x),$$

则称Y_n依分布收敛于Y，记作$Y_n \xrightarrow{d} Y$.

[1]对于一般的随机变量Y只要求对$F_Y(x)$的任意连续点x.

林德伯格–莱维(Lindberg–Lévy)中心极限定理

定理：设$\{X_n\}$为独立同分布随机变量序列，且方差期望存在，即$\mathbb{E}[X_n] = \mu$, $\mathrm{Var}[X_n] = \sigma^2$. 令

$$\xi_n := \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{\mathrm{Var}[\sum_{i=1}^n X_i]}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}.$$

则$\xi_n \xrightarrow{d} N(0,1)$, 即对任意$x \in \mathbb{R}$,

$$\lim_{n\to\infty} P(\xi_n \le x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2}\,\mathrm{d}t.$$

棣莫弗–拉普拉斯(De Moivre–Laplace)中心极限定理

定理：设$\{X_n\}$为独立同分布随机变量序列，同服从$B(1,p)$. 令

$$\xi_n := \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{\mathrm{Var}[\sum_{i=1}^n X_i]}} = \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}}.$$

则$\xi_n \xrightarrow{d} N(0,1)$, 即对任意$x \in \mathbb{R}$,

$$\lim_{n\to\infty} P(\xi_n \le x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2}\,\mathrm{d}t.$$

注1：棣莫弗–拉普拉斯中心极限定理是林德伯格–莱维中心极限定理的特例。
注2：更一般的中心极限定理参见林德伯格中心极限定理和李雅普诺夫中心极限定理，这两个定理不需要独立同分布的条件，只需满足相应的条件。
He, Z. & Zhu, L. Stat. Comput. (2017). Asymptotic normality of extensible grid sampling. 链接：`https://doi.org/10.1007/s11222-017-9794-y`

中心极限定理的应用

当n充分大时，可以认为

$$\xi_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \approx N(0,1)$$

或者

$$S_n = \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2)$$

或者

$$\frac{S_n}{n} \approx N(\mu, \sigma^2/n)$$

由此可以计算与S_n相关的事件的概率。比如

$$P(S_n \le x) \approx \Phi\left(\frac{x - n\mu}{\sqrt{n}\sigma}\right)$$

例1

假设报名学习某门选修课的学生人数是服从均值100的泊松分布的随机变量。学校领导决定，如果报名人数不少于120人，就分成两个班上。如果少于120人，就集中在一个班上。试问该选修班分成两个班上的概率是多少?

解：设报名人数为$X \sim Pois(100)$. 设$X_i \overset{\text{iid}}{\sim} Pois(1)$。由泊松分布的可加性得

$$X = \sum_{i=1}^{100} X_i$$

$\mathbb{E}[X] = \mathrm{Var}[X] = 100$. 由中心极限定理知，

$$P(X \ge 120) = 1 - \Phi((120-100)/\sqrt{100}) = 1 - \Phi(2) = 0.0228.$$

例2

例：设某地区原有一家小电影院，现拟筹建一所较大的电影院。根据分析，该地区每天平均看电影者约有$n = 1600$人，预计新电影院开业后，平均约有3/4的观众将去新电影院。现计划其座位数，要求座位数尽可能多，但“空座达到200或更多”的概率不能超过0.1，问设多少座位为好?
解：设座位数为m. 设

$$X_i = \begin{cases} 1, & \text{第}i\text{个观众去新院} \\ 0, & \text{其他}. \end{cases}$$

则$X_i \overset{\text{iid}}{\sim} B(1, 3/4)$. 令$S_n = X_1 + \cdots + X_n$. 依题意，

$$P(m - S_n \ge 200) = P(S_n \le m - 200) \le 0.1.$$

由中心极限定理得，

$$P(S_n \le m - 200) \approx \Phi((m - 200 - 3/4 * 1600)/\sqrt{1600 * (3/4)(1-3/4)}).$$

所以，$m \le \Phi^{-1}(0.1) * 10\sqrt{3} + 1400 \approx 1377.8$. 应设1377个座位。

例3

例：用调查对象中的收看比例k/n作为某电视节目的收视率p的估计。要有90%的把握，使k/n与p的差异不大于0.05，问至少要调查多少个对象？

解：设S_n表示n个调查对象中收看此节目的人数，则$S_n \sim B(n,p)$. 依题得，

$$P(|S_n/n - p| \leq 0.05) \geq 0.9.$$

由中心极限定理得

$$P(|S_n/n - p| \leq 0.05) = 2\Phi(0.05/\sqrt{p(1-p)/n}) - 1 \geq 0.9.$$

所以，

$$n \geq 400\Phi^{-1}(0.95)^2 p(1-p) \geq 100\Phi^{-1}(0.95)^2 = 100 * 1.645^2 = 270.6.$$

至少调查271个对象。

作业

第五章习题：1, 2, 4, 8, 10, 11

5月23日（周三）交！

概率论与数理统计

何志坚

华南理工大学数学学院

hezhijian@scut.edu.cn

第六章：数理统计的基本概念

数理统计是一门什么样的学科？

它使用概率论和其它数学方法，研究怎样收集（通过试验和观察）带有随机误差的数据，并在设定的模型（称为统计模型）之下，对这种数据进行分析（称为统计分析），以对所研究的问题作出推断（称为统计推断）。由于所收集的统计数据（资料）只能反映事物的局部特征，数理统计的任务就在于从统计资料所反映的局部特征以概率论作为理论基础去推断事物的整体特征。

- 数据收集
- 模型假定
- 数据分析/推断

本质：由局部（有限样本）推断整体（总体）

案例

- 数据收集：用n部手机进行测试，记录通话时间$X_1,\ldots,X_n$
- 模型假定：假设通话时间X服从正态分布$N(\mu,\sigma^2)$
- 数据分析：通过观测数据$x_1,\ldots,x_n$估计μ以及σ^2

目录

总体与样本

总体：研究对象X的概率分布，$X \sim F_X(\cdot, \theta)$，其中$\theta \in \Theta$为参数，$\Theta$为参数空间。

样本：从总体中随机抽取n个个体$X_1, X_2, \ldots, X_n$，称为总体的样本，n称为样本容量，简称样本量。样本具有二重性：

- 抽取之前无法预知它们的数值，因此$(X_1, \ldots, X_n)$为n维随机向量。
- 抽取后样本为具体的数，用小写字母$(x_1, \ldots, x_n)$表示，称为样本观测值。

简单随机样本：X_i相互独立，与X同分布。本课程只讨论简单随机样本。以后简称样本。

样本的分布

定义：设总体X具有分布函数$F(x)$，则样本$(X_1, \ldots, X_n)$的联合分布函数为

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i).$$

- 若X为离散型随机变量，分布列为$p_i = P(X = x_i)$, 则样本的联合分布列为
$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} p_i.$$
- 如X为连续型随机变量，密度函数为$p(x)$, 则样本的联合密度函数为
$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i).$$

正态样本的分布

例：如果$X \sim N(\mu, \sigma^2)$, 其样本$(X_1, \ldots, X_n)$的联合密度函数为：

$$\begin{aligned} f(x_1, \ldots, x_n) &= \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right). \end{aligned}$$

统计量

样本是总体的反映，但样本所含信息不能直接用于解决我们所要研究的问题，而需要把样本所含的信息进行数学上的加工使其浓缩起来，从而解决我们的问题。为此，数理统计学往往构造一个合适的依赖于样本的函数，我们称之为统计量。

定义：如果$(X_1, \ldots, X_n)$为来自总体的样本，若样本函数

$$T = T(X_1, \ldots, X_n)$$

中不含有任何未知参数，则称T为统计量。统计量的分布称为抽样分布。

例

例：设$X_1, \ldots, X_n$为来自$X \sim N(\mu, \sigma^2)$的样本, 若μ已知，σ未知，判断以下样本函数是否为统计量

- $$T_1 = \frac{\sqrt{n}(\sum_{i=1}^{n} X_i - \mu)}{\sigma}$$
- $$T_2 = \min(X_1, \ldots, X_n)$$

常见的统计量

设$X_1, \ldots, X_n$为来自X的样本。

- 样本均值：
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- 样本方差：
$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
- 修正样本方差：
$$S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}S_n^2$$
- 样本标准差：
$$S_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

常见的统计量

设$X_1, \ldots, X_n$为来自X的样本。

- 样本k阶原点矩：
$$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$
- 样本k阶中心矩：
$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$
- 顺序统计量：
$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$
其中$X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$, $X_{(k)}$为$X_1, \ldots, X_n$的递增排序的第k位。$X_{(n)} - X_{(1)}$称为样本极差。
- 样本中位数：
$$\tilde{X} = \begin{cases} X_{(\frac{n+1}{2})}, & n\text{为奇数} \\ (X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)})/2, & n\text{为偶数} \end{cases}$$

目录

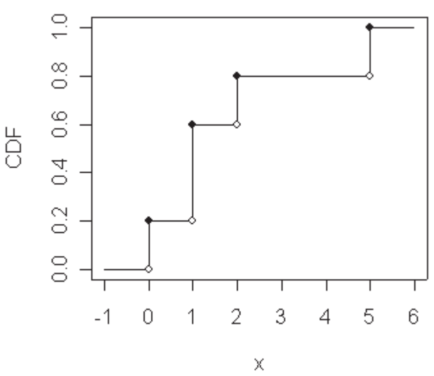
经验分布函数

目标：通过样本的观测值构造一种函数来近似总体的分布函数

定义：设总体X的样本$(X_1, \ldots, X_n)$的一次观测值$(x_1, \ldots, x_n)$，并将它们由小到大排列$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$，经验分布函数(或称样本分布函数)定义为

$$F_n^X(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{x_i \le x\} = \begin{cases} 0, & x < x_{(1)} \\ 1/n, & x_{(1)} \le x < x_{(2)} \\ 2/n, & x_{(2)} \le x < x_{(3)} \\ \vdots & \\ k/n, & x_{(k)} \le x < x_{(k+1)} \\ \vdots & \\ 1, & x > x_{(n)} \end{cases}.$$

经验分布函数示意图



经验分布函数的性质

固定的x和n，$F_n^X(x)$表示事件$\{X \le x\}$的频率，由强大数定律知，

$$F_n^X(x) \xrightarrow{\text{a.s}} P(X \le x) = F_X(x),$$

即

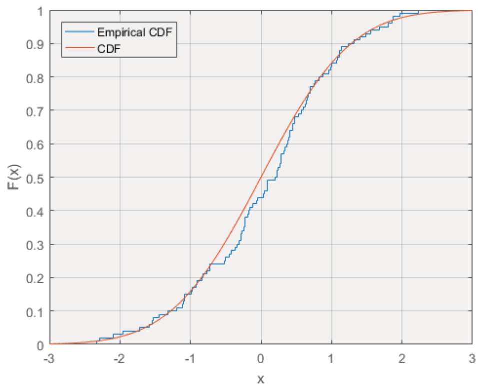
$$P\left(\lim_{n\to\infty} F_n^X(x) = F_X(x)\right) = 1.$$

格里汶科定理给出更强的结果（几乎处处一致收敛）：

$$P\left(\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F_n^X(x) - F_X(x)| = 0\right) = 1.$$

注：由此可见，当n相当大时，经验分布函数$F_n^X(x)$是母体分布函数$F_X(x)$的一个良好近似。数理统计学中一切以样本为依据，其理由就在于此。

经验分布函数示意图



目录

抽样分布

统计量的分布称为抽样分布。样本均值和样本方差的数字特征：

定理：设$X_1,\dots,X_n$为来自X的样本，$\mathbb{E}[X]=\mu$，$\mathrm{Var}[X]=\sigma^2$，则

- 样本均值的期望和方差分别为

$$\mathbb{E}[\bar{X}]=\mu,\ \mathrm{Var}[\bar{X}]=\frac{\sigma^2}{n}$$

- 样本方差和修正样本方差的期望分别为

$$\mathbb{E}[S_n^2]=\frac{n-1}{n}\sigma^2,\ \mathbb{E}[S_n^{*2}]=\sigma^2.$$

证：

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

$$\mathbb{E}[S_n^2]=\mathbb{E}[X^2]-\mathbb{E}[\bar{X}^2]=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

三种重要的概率分布

1. χ^2分布
2. t分布
3. F分布

χ^2分布

定义：设$X_i \overset{\text{iid}}{\sim} N(0,1), i=1,\dots,n$，则称随机变量

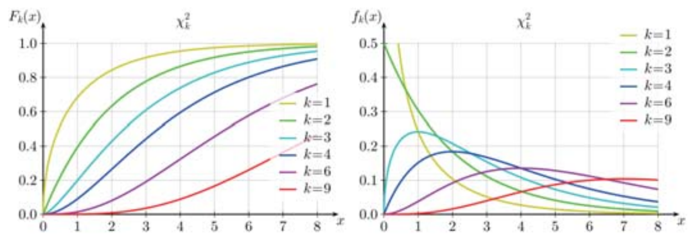
$$X=X_1^2+\cdots+X_n^2$$

服从自由度为n的卡方分布，记为$X\sim\chi^2(n)$(或者写成χ_n^2)。其密度函数为

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

其中$\Gamma(s)=\int_0^\infty x^{s-1}e^{-x}\,\mathrm{d}x(s>0)$。容易发现$\Gamma(n)=(n-1)!$.

χ^2分布的分布函数与密度函数曲线



χ^2分布的性质

定理：如果$X \sim \chi^2(n)$，则

$$\mathbb{E}[X] = n,\ \mathrm{Var}[X] = 2n.$$

如果$X \sim \chi^2(n)$，$Y \sim \chi^2(m)$且它们独立，则

$$X + Y \sim \chi^2(n+m)(\text{可加性}).$$

注： 英国统计学家费歇（R.A.Fisher）曾证明，当$n \to \infty$时，

$$\frac{X-n}{\sqrt{2n}} \xrightarrow{\mathrm{d}} N(0,1).$$

这就是中心极限定理的应用。

学生t分布

定义：设$X \sim N(0,1)$，$Y \sim \chi^2(n)$，且它们独立，则称随机变量

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

服从自由度为n的学生t分布（简称t分布），记为$T \sim t(n)$. 其密度函数为

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x \in \mathbb{R}.$$

休息一会儿 51

休息一会儿

小故事：t检验、啤酒、"学生"与威廉·戈瑟特

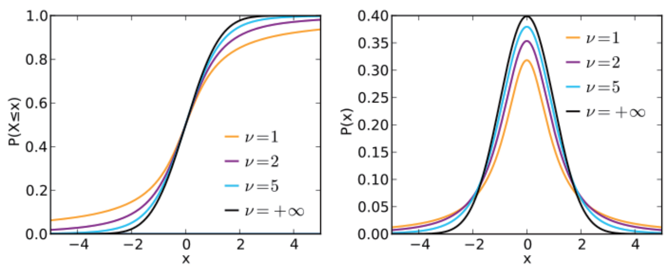
1954 年该厂开始出版《吉尼斯世界纪录大全》.

1899 年，由于爱尔兰都柏林的吉尼斯啤酒厂热衷于聘用剑桥、牛津的优秀毕业生，学化学的牛津毕业生威廉·戈瑟特 (William Gosset, 1876—1937) 到该厂就职，希望将他的生物化学知识用于啤酒生产过程. 为降低啤酒质量监控的成本，戈瑟特发明了t检验法，1908 年在 *Biometrika* 发表. 为防止泄漏商业机密，戈瑟特发表文章时用了笔名"学生"，于是该方法被称为"学生氏t检验" (Student's t-test).

吉尼斯啤酒厂是一家很有远见的企业，为保持技术人员的高水准，该厂像高校一样给予技术人员"学术假"，1906—1907 年戈瑟特得以到"统计学之父"卡尔·皮尔逊 (Karl Pearson, 1857—1936) 教授在伦敦大学学院 (University College London, 简称 UCL) 的实验室访问学习. 因此，很难说t检验法是戈瑟特在啤酒厂还是在 UCL 访学期间提出的，但"学生"与戈瑟特之间的联系是被 UCL 的统计学家们发现的，尤其因为皮尔逊教授恰是 *Biometrika* 的主编.

推荐阅读《女士品茶》

t分布的分布函数与密度函数曲线



注： 密度函数是偶函数

t分布的性质

设$X \sim t(n)$.

- 当$n=1$时，t分布成为柯西分布.

$$f(x) = \frac{\Gamma(1)}{\sqrt{\pi}\Gamma\left(\frac{1}{2}\right)}\left(1+x^2\right)^{-1},\ x \in \mathbb{R}.$$

  此时，方差和期望都不存在。
- 当$n \geq 2$时，$\mathbb{E}[X]=0$. 对于k阶矩，$\mathbb{E}[X^k]$存在当且仅当$k < n$.
- 当$n \geq 3$时，t分布的方差才存在，即

$$\mathrm{Var}[X] = \frac{n}{n-2}.$$

-

$$\lim_{n\to\infty} f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

  当$n > 45$时,可以认为t分布与$N(0,1)$接近。

例

例：设X, Y这两个总体相互独立，$X \sim N(0,16)$，$Y \sim N(0,9)$，$X_1,\ldots,X_9$为X的样本，$Y_1,\ldots,Y_{16}$为Y的样本，求统计量

$$Z = \frac{X_1+\cdots+X_9}{\sqrt{Y_1^2+\cdots+Y_{16}^2}}$$

所服从的分布。

证：$X_1+\cdots+X_9 \sim N(0,12^2)$，所以$(X_1+\cdots+X_9)/12 \sim N(0,1)$. 又因为，$(Y_1^2+\cdots+Y_{16}^2)/9 \sim \chi^2(16)$. 所以，

$$Z = \frac{(X_1+\cdots+X_9)/12}{\sqrt{\frac{(Y_1^2+\cdots+Y_{16}^2)/9}{16}}} \sim t(16)$$

F分布

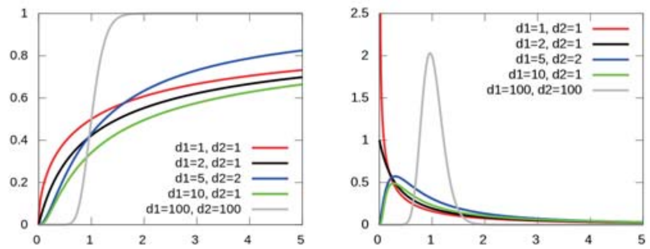
定义：设$X \sim \chi^2(m), Y \sim \chi^2(n)$，且$X, Y$相互独立，则随机变量

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

称为服从第一自由度为m、第二自由度为n的F分布，记$Z \sim F(m, n)$. 其密度函数为

$$f(x) = \begin{cases} \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1}(1+mx/n)^{-(m+n)/2}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

F分布的分布函数与密度函数曲线



F分布的性质

- 如果$Z \sim F(m, n)$, 则$1/Z \sim F(n, m)$.
- 如果$T \sim t(n)$, 则$T^2 \sim F(1, n)$.

分位数

定义：对于任意$\alpha \in (0, 1)$, α分位数x_α满足$F(x_\alpha) = \alpha$.

- 标准正态分布分位数记为u_α, 满足$P(|Z| \le u_{1-\alpha/2}) = 1 - \alpha$
- t分布分位数记为$t_\alpha(n)$, 满足$P(|T| \le t_{1-\alpha/2}(n)) = 1 - \alpha$
- χ^2分布分位数记为$\chi^2_\alpha(n)$
- F分布分位数记为$F_\alpha(m, n)$

注1：在分位点表中对于标准正态分布、t分布和F分布只能查到$\alpha > 1/2$的分位数，需利用以下对称性间接查$\alpha < 1/2$的分位数：

$$u_\alpha = -u_{1-\alpha},\ t_\alpha(n) = -t_{1-\alpha}(n),\ F_\alpha(m, n) = \frac{1}{F_{1-\alpha}(n, m)}$$

注2：对于$t(n)$分布，由于当$n \to \infty$时，其极限分布为$N(0, 1)$, 所以自由度n比较大时，$t_\alpha(n) \approx u_\alpha$.

注3：若$X \sim \chi^2(n)$分布，由于当$n \to \infty$时，$(X - n)/\sqrt{2n} \xrightarrow{d} N(0, 1)$, 所以自由度$n$比较大时，$\chi^2_\alpha(n) \approx u_\alpha\sqrt{2n} + n$.

分位数

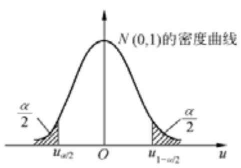


图 6.3.4. $N(0,1)$ 分布分位点示意图.

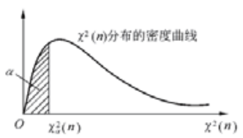


图 6.3.5. $\chi^2(n)$ 分布分位点示意图.

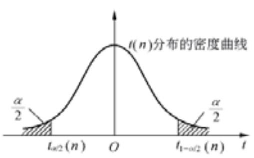


图 6.3.6. $t(n)$ 分布分位点示意图.

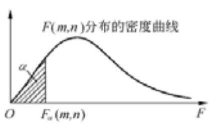


图 6.3.7. $F(m,n)$ 分布分位点示意图.

例

例：查表求以下分位数：

- $\chi^2_{0.99}(10)$，$\chi^2_{0.05}(10)$和$\chi^2_{0.95}(60)$
- $t_{0.95}(10)$，$t_{0.1}(10)$和$t_{0.9}(50)$
- $F_{0.99}(5, 4)$，$F_{0.05}(3, 7)$

解：

$$\chi^2_{0.99}(10) = 23.209,\ \chi^2_{0.05}(20) = 10.851,\ \chi^2_{0.95}(60) = u_{0.95}\sqrt{2 \times 60} + 60 = 78.018$$

$$t_{0.95}(10) = 1.8125,\ t_{0.1}(10) = -t_{0.9}(10) = -1.3722\ t_{0.9}(50) = u_{0.9} = 1.28$$

$$F_{0.99}(5, 4) = 5.5,\ F_{0.05}(3, 7) = 1/F_{0.95}(7, 3) = 1/8.89 = 0.112$$

单个正态总体的抽样分布

设$(X_1,\dots,X_n)$为来自总体$X\sim N(\mu,\sigma^2)$的样本，则有

$$\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}=\sum_{i=1}^n\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$$

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

注意到

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

$$\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}=\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{\sigma^2}+\frac{(\bar{X}-\mu)^2}{\sigma^2/n}$$

单个正态总体的抽样分布

定理：设$(X_1,\dots,X_n)$为来自总体$X\sim N(\mu,\sigma^2)$的样本，则

- 样本均值$\bar{X}$与样本方差S_n^2相互独立，
- $\bar{X}\sim N(\mu,\sigma^2/n)$，
- $$\frac{nS_n^2}{\sigma^2}=\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{\sigma^2}\sim\chi^2(n-1).$$

证：由正态分布的可加性得$\bar{X}\sim N(\mu,\sigma^2/n)$. 下证独立性，令$Y_i=X_i-\bar{X}$. 则有$\sum_{i=1}^n Y_i=0$. 所以，$S_n^2$为$Y_1,\dots,Y_{n-1}$的函数。由于$(\bar{X},Y_1,\dots,Y_{n-1})$服从$n$维正态分布，所以只需验证$\mathrm{Cov}(\bar{X},Y_i)=\mathrm{Cov}(\bar{X},X_i-\bar{X})=0$即可。

当$n=2$时，

$$\frac{nS_n^2}{\sigma^2}=\left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right)^2\sim\chi^2(1)$$

假设$n=k$时结论成立，当$n=k+1$时，

$$\frac{(k+1)S_{k+1}^2}{\sigma^2}=\frac{kS_k^2}{\sigma^2}+\frac{k(X_{k+1}-\bar{X}_k)^2}{(k+1)\sigma^2}=\chi^2(k-1)+\chi^2(1)=\chi^2(k)$$

推论一：标准正态总体

推论：设$(X_1,\dots,X_n)$为来自总体$X\sim N(0,1)$的样本，则

- 样本均值$\bar{X}$与样本方差S_n^2相互独立，
- $$\bar{X}\sim N(0,\frac{1}{n}),$$
- $$nS_n^2=\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{\sigma^2}\sim\chi^2(n-1).$$

推论二

推论：设$(X_1,\dots,X_n)$为来自总体$X\sim N(\mu,\sigma^2)$的样本，则

$$\frac{\sqrt{n-1}(\bar{X}-\mu)}{S_n}=\frac{\sqrt{n}(\bar{X}-\mu)}{S_n^*}\sim t(n-1)$$

证明:令

$$Y=\frac{\bar{X}-\mu}{\sqrt{\mathrm{Var}[\bar{X}]}}=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$$

因为，$\bar{X}$与S_n^2独立，所以Y与$Z=\frac{nS_n^2}{\sigma^2}$独立。由于$Z\sim\chi^2(n-1)$, 所以

$$\frac{Y}{\sqrt{Z/(n-1)}}=\frac{\sqrt{n-1}(\bar{X}-\mu)}{S_n}\sim t(n-1)$$

例

例：设总体$X\sim N(\mu,4)$, 有样本$X_1,\dots,X_n$, 当样本容量n至少多大时，才使得$P(|\bar{X}-\mu|\le 0.01)\ge 0.95$?

解：$\bar{X}\sim N(\mu,4/n)$. 所以$Z=(\bar{X}-\mu)/(2/\sqrt{n})\sim N(0,1)$.

$$P(|\bar{X}-\mu|\le 0.01)=P(|Z|\le 0.05\sqrt{n})\ge 0.95.$$

所以，$0.05\sqrt{n}\ge u_{0.975}$，$n\ge 400u_{0.975}^2=400\times 1.96^2=1536.6\approx 1537$.

两个正态总体下的抽样分布

定理：设总体$X\sim N(\mu_1,\sigma_1^2)$, $(X_1,\dots,X_m)$为其样本，样本均值为$\bar{X}$, 样本方差为S_{1m}^2; 另有与X独立的总体$Y\sim N(\mu_2,\sigma_2^2)$, $(Y_1,\dots,Y_n)$为其样本，样本均值为$\bar{Y}$, 样本方差为S_{2n}^2. 则

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1).$$

$$\frac{mS_{1m}^2/\sigma_1^2/(m-1)}{nS_{2n}^2/\sigma_2^2/(n-1)}=\frac{S_{1m}^{*2}\sigma_2^2}{S_{2n}^{*2}\sigma_1^2}\sim F(m-1,n-1).$$

证明：利用单个正态整体下抽样分布的结论：$\bar{X}\sim N(\mu_1,\sigma_1^2/m)$, $\bar{Y}\sim N(\mu_2,\sigma_2^2/n)$, 且$\bar{X},\bar{Y}$独立，所以，$\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2,\sigma_1^2/m+\sigma_2^2/n)$. 又$mS_{1m}^2/\sigma_1^2\sim\chi^2(m-1)$, $nS_{2n}^2/\sigma_2^2\sim\chi^2(n-1)$, 且两者独立，根据$F$分布的定义可得第二个结论。

两个正态总体（相同方差）下的抽样分布

定理：设总体$X \sim N(\mu_1, \sigma^2)$, $(X_1, \ldots, X_m)$为其样本，样本均值为$\bar{X}$, 样本方差为S_{1m}^2; 另有与X独立的总体$Y \sim N(\mu_2, \sigma^2)$, $(Y_1, \ldots, Y_n)$为其样本，样本均值为$\bar{Y}$, 样本方差为S_{2n}^2. 则

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{1/m + 1/n}} \sim t(m + n - 2),$$

其中$S_w = \sqrt{(mS_{1m}^2 + nS_{2n}^2)/(m + n - 2)}$.

证明：令

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2/m + \sigma^2/n}} \sim N(0, 1)$$

因为$mS_{1m}^2/\sigma^2 \sim \chi^2(m-1)$, $nS_{2n}^2/\sigma^2 \sim \chi^2(n-1)$ 且它们独立，所以

$$V = mS_{1m}^2/\sigma^2 + nS_{2n}^2/\sigma^2 \sim \chi^2(n + m - 2).$$

又U, V独立，所以$U/\sqrt{V/(m + n - 2)} \sim t(m + n - 2)$.

例1：$\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$的值已知

例：设总体$X \sim N(6, 1)$, $Y \sim N(5, 1)$有$m = n = 10$两个独立样本，求$\bar{X} - \bar{Y}$小于1.3的概率。

解： 已知

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} = \frac{(\bar{X} - \bar{Y}) - 1}{\sqrt{1/10 + 1/10}} \sim N(0, 1)$$

所以，

$$P(\bar{X} - \bar{Y} < 1.3) = P(Z < (1.3 - 1)/\sqrt{1/5}) = P(Z < 0.67) = \Phi(0.67) = 0.7486.$$

例2：μ_1, μ_2的值已知，且$\sigma_1^2 = \sigma_2^2$

例：设总体$X \sim N(6, \sigma^2)$, $Y \sim N(5, \sigma^2)$有$m = n = 10$两个独立样本，$S_{1m}^{*2} = 0.9130$, $S_{2n}^{*2} = 0.9816$, 求$\bar{X} - \bar{Y}$小于1.3的概率。

解： 已知

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{1/m + 1/n}} \sim t(m + n - 2) = t(18),$$

其中$S_w = \sqrt{((m-1)S_{1m}^{*2} + (n-1)S_{2n}^{*2})/(m + n - 2)} = 0.9733$. 所以，

$$P(\bar{X} - \bar{Y} < 1.3) = P(Z < (1.3 - 1)/(\sqrt{1/5} \times 0.9733)) = P(Z < 0.6884) = 0.74.$$

例3：σ_1^2, σ_2^2的值已知，但μ_1, μ_2的值未知

例：设总体$X \sim N(\mu, 3)$, $Y \sim N(\mu, 5)$有$m = 10$, $n = 15$两个独立样本，求两个修正样本方差之比S_{1m}^{*2}/S_{2n}^{*2}大于1.272的概率。

解： 已知

$$Z = \frac{S_{1m}^{*2}\sigma_2^2}{S_{2n}^{*2}\sigma_1^2} \sim F(m-1, n-1) = F(9, 14)$$

所以，

$$P(S_{1m}^{*2}/S_{2n}^{*2} > 1.272) = P(Z > 1.272 \times 5/3) = P(Z > 2.12) = 0.1.$$

顺序统计量分布

定理：若$X_1, \ldots, X_n$独立同分布，分布函数和密度函数分别为$F_X(x), f_X(x)$. 则$X_{(1)} = \min(X_1, \ldots, X_n)$的分布函数和密度函数分别

$$\begin{cases} F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n \\ f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_X(x). \end{cases}$$

$X_{(n)} = \max(X_1, \ldots, X_n)$的分布函数和密度函数分别

$$\begin{cases} F_{X_{(n)}}(x) = F_X(x)^n \\ f_{X_{(n)}}(x) = nF_X(x)^{n-1} f_X(x). \end{cases}$$

更一般地，

$$f_{X_{(k)}}(x) = \frac{n!}{(n-k)!(k-1)!} F_X(x)^{k-1}(1 - F_x(x))^{n-k} f_X(x), k = 1, \ldots, n.$$

作业

第六章习题：2, 3, 4, 8, 11, 13, 17, 18

6月6日（周三）交！

概率论与数理统计

何志坚

华南理工大学数学学院

hezhijian@scut.edu.cn

第七章：参数估计

参数估计

在实际问题中，对于一个总体X往往是仅知其分布的类型，而其中所含的一个或几个参数的值却是未知的，因此只有在确定这些参数后，才能通过其分布来计算概率，如何确定这些参数的数值呢？这就是统计推断中的“参数估计”问题。

目录

参数的点估计

定义：构造一个统计量$\hat\theta$对参数θ作定值的估计称为**参数的点估计**

$$\begin{cases} \text{点估计量：} \hat\theta &= T(X_1,\dots,X_n) \\ \text{点估计值：} \hat\theta &= T(x_1,\dots,x_n) \end{cases}$$

矩法估计

矩估计的想法来源于大数定理。如果总体X存在k阶矩，对任意$\epsilon>0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i^k - \mathbb{E}[X^k]\right| \ge \epsilon\right) = 0.$$

这说明，当样本容量n较大时，样本k阶矩与总体k阶矩差别很小。**矩法估计就是用样本k阶矩估计总体的k阶矩**。通常用$\hat\theta_M$表示。一般步骤如下：

- **第一步**：列出估计式
$$E[X^k] = g_k(\theta_1,\dots,\theta_m),\ k=1,\dots,m.$$
- **第二步**：求解关于估计量的方程组
$$\theta_k = \theta_k(E[X^1],\dots,E[X^m]),\ k=1,\dots,m.$$
- **第三步**：用$M_k=\frac{1}{n}\sum_{i=1}^n X_i^k$替代$E[X^k]$得到矩估计
$$\hat\theta_k = \theta_k(M_1,\dots,M_m),\ k=1,\dots,m.$$

例1

例：求总体X的期望$\mu=\mathbb{E}[X]$与方差$\sigma^2=\mathrm{Var}[X]$的矩估计。

解：(1)列出估计式

$$\begin{cases}\mathbb{E}[X] & =\mu\\ \mathbb{E}[X^2] & =\mathrm{Var}[X]+(\mathbb{E}[X])^2=\mu^2+\sigma^2\end{cases}$$

(2)求解关于估计量的方程组

$$\begin{cases}\mu & =\mathbb{E}[X]\\ \sigma^2 & =\mathbb{E}[X^2]-(\mathbb{E}[X])^2\end{cases}$$

所以，$\hat{\mu}_M=\bar{X}$, $\hat{\sigma}_M^2=\frac{1}{n}\sum_{i=1}^n X_i^2-(\bar{X})^2=S_n^2$.
注：不难证明，总体的各阶中心矩的矩估计就是样本各阶中心矩。

例2

例：设总体$X\sim U[a,b]$, 求a,b的矩估计。

解：$\mathbb{E}[X]=(a+b)/2$, $\mathrm{Var}[X]=(b-a)^2/12$. 所以，

$$\begin{cases}a & =\mathbb{E}[X]-\sqrt{3\mathrm{Var}[X]}\\ b & =\mathbb{E}[X]+\sqrt{3\mathrm{Var}[X]}.\end{cases}$$

$$\begin{cases}\hat{a}_M & =\bar{X}-\sqrt{3}S_n\\ \hat{b}_M & =\bar{X}+\sqrt{3}S_n.\end{cases}$$

例

例：设总体X的分布密度为

$$f(x)=\frac{\theta}{2}e^{-\theta|x|},\ x\in\mathbb{R},\theta>0.$$

求θ的矩估计。

解：

$$\mathbb{E}[X]=0,\ \mathbb{E}[X^2]=\int_{-\infty}^{\infty}x^2\frac{\theta}{2}e^{-\theta|x|}\,\mathrm{d}x=\theta\int_0^{\infty}x^2e^{-\theta x}\,\mathrm{d}x=\frac{2}{\theta^2}$$

$$\hat{\theta}_M=\sqrt{\frac{2n}{\sum_{i=1}^n X_i^2}}.$$

除外，还可以由$\mathbb{E}[|X|]=1/\theta$得到另一种矩估计。

最大似然估计

最大似然估计法是求估计的另一种方法。它最早由高斯(C.F.Gauss)提出，后来被费歇(R. A. Fisher)完善。最大似然估计这一名称也是费歇给的。这是一个目前仍得到广泛应用的方法。它是建立在最大似然原理基础上的一个统计方法。

例：设有外形完全相同的两个箱子，甲箱中有99个白球和1个黑球，乙箱中有99个黑球和1个白球，今随机地抽取一箱并从中随机抽取一球，结果取得白球，问这球是从哪个箱子中取出?

最大似然原理：最先出现的是概率最大的

最大似然估计

似然函数：样本$(X_1,\ldots,X_n)$的联合分布（密度）函数

$$L(x_1,\ldots,x_n;\theta)=L(\theta)=\begin{cases}\prod_{i=1}^n P(X_i=x_i;\theta), & \text{离散型}\\ \prod_{i=1}^n f(x_i;\theta), & \text{连续型.}\end{cases}$$

给定样本观测值$(x_1,\ldots,x_n)$, 记$L(x_1,\ldots,x_n;\theta)$的最大值点为$\theta=T(x_1,\ldots,x_n)$. 则θ的最大似然估计量(MLE, maximum likelihood estimator)为

$$\hat{\theta}_L=T(X_1,\ldots,X_n).$$

最大似然估计的一般步骤

第一步：写出似然函数$L(x_1,\ldots,x_n;\theta)$
第二步：若似然函数L是θ的可微函数，则最大值必然满足似然方程

$$\frac{\mathrm{d}L}{\mathrm{d}\theta}=0$$

解出θ, 并验证其是否是极大值：

$$\frac{\mathrm{d}^2L}{\mathrm{d}\theta^2}<0.$$

注1：为方便求导，一般求对数似然函数求极大值点

$$\ln L(x_1,\ldots,x_n;\theta)=\begin{cases}\sum_{i=1}^n \ln P(X_i=x_i;\theta), & \text{离散型}\\ \sum_{i=1}^n \ln f(x_i;\theta), & \text{连续型.}\end{cases}$$

注2：若有多个参数$\theta_1,\ldots,\theta_m$，对每个变量求偏导，联立m个方程求解。

例1：0-1离散型

例：设总体$X \sim B(1,p)$, 从中抽取样本$X_1,\dots,X_n$的观测值为$x_1,\dots,x_n$. 求参数p的最大似然估计。

解: 似然函数为

$$L(x_1,\dots,x_n;p)=\prod_{i=1}^{n}P(X=x_i)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

令$y=\sum_{i=1}^n x_i$, 对数似然函数为:

$$\ln L = y\ln p + (n-y)\ln(1-p).$$

对数似然方程为:

$$\frac{\mathrm{d}\ln L}{\mathrm{d}p}=y/p-(n-y)/(1-p)=0.$$

解得$p=y/n=\frac{1}{n}\sum_{i=1}^n x_i$. 因为$\frac{\mathrm{d}^2\ln L}{\mathrm{d}p^2}<0$, 所以$p=y/n$是极大值。$\hat{p}_L=\bar{X}$.

例2：正态分布

例：设总体$X \sim N(\mu,\sigma^2)$, 从中抽取样本$X_1,\dots,X_n$的观测值为$x_1,\dots,x_n$. 求参数μ,σ^2的最大似然估计。

解: 似然函数为

$$L(x_1,\dots,x_n;\mu,\sigma^2)=\prod_{i=1}^{n}f(x_i)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu)^2/(2\sigma^2)}$$

令$\theta_1=\mu,\theta_2=\sigma^2$, 对数似然函数为:

$$\ln L=(n/2)\ln(2\pi)-(n/2)\ln\theta_2-\frac{\sum_{i=1}^n(x_i-\theta_1)^2}{2\theta_2}$$

对数似然方程组为:

$$\begin{cases}\frac{\partial \ln L}{\partial\theta_1} & =\frac{\sum_{i=1}^n(x_i-\theta_1)}{\theta_2}=0\\ \frac{\partial \ln L}{\partial\theta_2} & =-\frac{n}{2\theta_2}+\frac{\sum_{i=1}^n(x_i-\theta_1)^2}{2\theta_2^2}=0\end{cases}$$

解得$\hat{\mu}_L=\bar{X}$, $\hat{\sigma}_L^2=S_n^2$. (可以验证二阶导函数非正定，即取得极大值。)

例3

例：设总体$X \sim U[a,b]$, 从中抽取样本$X_1,\dots,X_n$的观测值为$x_1,\dots,x_n$. 求参数a,b的最大似然估计。

解: 似然函数为

$$L(x_1,\dots,x_n;a,b)=\prod_{i=1}^{n}\frac{1}{b-a}\mathbf{1}\{a\le x_i\le b\}=\frac{1}{(b-a)^n}\prod_{i=1}^{n}\mathbf{1}\{a\le x_i\le b\}$$

注意到L关于a,b不可微。容易观察到，
当$a=\min_{i=1,\dots,n}\{x_i\}$, $b=\max_{i=1,\dots,n}\{x_i\}$时$L$取得最大值。故

$$\hat{a}_L=X_{(1)},\ \hat{b}_L=X_{(n)}.$$

关于最大似然估计的一些说明

- 最大似然估计的不变性：如果$\hat{\theta}$是θ的最大似然估计，则对任一函数$g(\theta)$, 其最大似然估计为$g(\hat{\theta})$.
- 当分布中有多余的参数或者数据为截尾或缺失时，似然函数的求极大值比较困难。针对这种问题，文献 Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". **Journal of the Royal Statistical Society, Series B**. 39 (1): 1–38. (cited by 53600, 2018/5/16) 提出了一种有效的Expectation–Maximization (EM)算法。

常见分布的矩估计与最大似然估计的对比

分布名称	记号	期望	方差	矩估计	极大似然
0-1分布	$B(1,p)$	p	pq	$\hat{p}_M=\bar{X}$	$\hat{p}_L=\bar{X}$
泊松分布	$Pois(\lambda)$	λ	λ	$\hat{\lambda}_M=\bar{X}$	$\hat{\lambda}_L=\bar{X}$
几何分布	$Geo(p)$	$1/p$	q/p^2	$\hat{p}_M=1/\bar{X}$	$\hat{p}_L=1/\bar{X}$
均匀分布	$U[a,b]$	$(a+b)/2$	$(b-a)^2/12$	$\hat{a}_M=\bar{X}-\sqrt{3}S_n$ $\hat{b}_M=\bar{X}+\sqrt{3}S_n$	$\hat{a}_L=X_{(1)}$ $\hat{b}_L=X_{(n)}$
指数分布	$Exp(\lambda)$	$1/\lambda$	$1/\lambda^2$	$\hat{\lambda}_M=1/\bar{X}$	$\hat{\lambda}_L=1/\bar{X}$
正态分布	$N(\mu,\sigma^2)$	μ	σ^2	$\hat{\mu}_M=\bar{X}$ $\hat{\sigma}_M^2=S_n^2$	$\hat{\mu}_L=\bar{X}$ $\hat{\sigma}_L^2=S_n^2$

练习：计算上述前六种分布参数的矩估计与最大似然估计。

顺序统计量估计

总体是连续型随机变量且分布密度对称时，总体中位数就是均值。此时可用样本中位数估计总体均值μ，用样本极差估计总体标准差σ, 即

$$\begin{cases}\hat{\mu} & =\tilde{X}\\ \hat{\sigma} & =X_{(n)}-X_{(1)}.\end{cases}$$

三种点估计方法的比较

- 矩估计法（也称数字特征法）
  直观意义比较明显，但要求总体k阶矩存在。
- 极大似然估计法
  具有一些理论上的优点，但要求似然函数可微。
- 顺序统计量法
  使用起来方便，无需多大计算，但准确度不高。

目录

估计量优劣性的评价

定义：设总体$X \sim F(x;\theta)$，$\theta \in \Theta$，统计量$T(X_1,\dots,X_n)$为$g(\theta)$的估计量。

- 无偏估计量：$\mathbb{E}[T(X_1,\dots,X_n)] = g(\theta)$
- 渐近无偏估计量：$\lim_{n\to\infty} \mathbb{E}[T(X_1,\dots,X_n)] = g(\theta)$

注：无偏性意味着：虽然估计量T由于随机可能偏离真值$g(\theta)$，但取其平均值（期望）却等于$g(\theta)$. 即没有系统偏差。

例1

例：样本均值是总体的均值的无偏估计；样本方差是总体方差的渐近无偏估计；修正样本方差是总体方差的无偏估计。

解：

$$\mathbb{E}[\bar{X}] = \mathbb{E}[(X_1+\cdots+X_n)/n] = \mathbb{E}[X]$$
$$\mathbb{E}[S_n^2] = \frac{n-1}{n}\mathrm{Var}[X] \to \mathrm{Var}[X]$$
$$\mathbb{E}[S_n^{*2}] = \mathrm{Var}[X]$$

例2

例：设总体X的期望μ方差为σ^2, $X_1,\dots,X_n$为其样本，证明下列估计量

$$\hat{\mu} = \sum_{i=1} C_i X_i$$

为μ的无偏估计的充要条件是$\sum_{i=1}^n C_i = 1$. 在满足该条件前提下，C_i取何值时，$\hat{\mu}$的方差最小。

解：

$$\mathbb{E}[\hat{\mu}] = \mu \Leftrightarrow \sum_{i=1}^n C_i = 1$$
$$\mathrm{Var}[\hat{\mu}] = \sigma^2 \sum_{i=1}^n C_i^2 \geq \sigma^2 \frac{(C_1+\cdots+C_n)^2}{n} = \frac{\sigma^2}{n}.$$

而且唯一的最小值在$C_i = 1/n, i = 1,\dots,n$处取得。

一致最小方差无偏估计量

从上一个例子中可以看出，无偏估计量不唯一。由于方差是度量分布的离散程度，一个好的估计量不仅应该是待估参数的无偏估计，而且应该有尽可能小的方差。

定义1：如果$T_1(X_1,\dots,X_n)$和$T_2(X_1,\dots,X_n)$均为$g(\theta)$的无偏估计，如果

$$\mathrm{Var}[T_1] \leq \mathrm{Var}[T_2],$$

则称T_1比T_2更有效。

注：对于上一个例子，样本均值$\bar{X}$比其他的估计量更有效。

定义2：如果$T_0(X_1,\dots,X_n)$为$g(\theta)$的无偏估计，如果对于$g(\theta)$的任意无偏估计量$T(X_1,\dots,X_n)$都有

$$\mathrm{Var}[T_0] \leq \mathrm{Var}[T],\ \forall \theta \in \Theta$$

则称T_0为$g(\theta)$的一致最小方差无偏估计量。

相合估计量

我们不仅希望一个估计是无偏的，且具有较小的方差，有时还希望当子样容量无限增大时，即观察次数无限增多时，估计能在某种意义下越来越接近被估计的参数的真实值，这就是所谓一致性的要求。

定义：设总体$X \sim F(x;\theta)$，$\theta \in \Theta$，统计量$T(X_1,\dots,X_n)$为$g(\theta)$的估计量。如果$T(X_1,\dots,X_n) \xrightarrow{P} g(\theta)$，则称$T(X_1,\dots,X_n)$为$g(\theta)$的相合估计量（一致估计量），即对于任意$\epsilon > 0$有，

$$\lim_{n\to\infty} P(|T(X_1,\dots,X_n) - g(\theta)| \geq \epsilon) = 0,$$

或者等价地

$$\lim_{n\to\infty} P(|T(X_1,\dots,X_n) - g(\theta)| < \epsilon) = 1.$$

例

例：设总体X的期望μ方差为σ^2，$X_1,\dots,X_n$为其样本，证明

- 样本均值$\bar{X}$是μ的相合估计量；
- 样本k阶原点矩M_k是总体k阶原点矩$\mathbb{E}[X^k]$的相合估计量；
- 样本方差S_n^2和修正样本方差S_n^{2*}都是σ^2的相合估计量。

证明：由辛钦大数定律知，$\bar{X} \xrightarrow{P} \mu$，$M_k \xrightarrow{P} \mathbb{E}[X^k]$.

$$S_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 \xrightarrow{P} \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2$$

同理，

$$S_n^{2*} = \frac{n-1}{n} S_n^2 \xrightarrow{P} \sigma^2$$

注：这里用到依概率收敛的性质：假设$X_n \xrightarrow{P} X$，$Y_n \xrightarrow{P} Y$.
则$X_n + Y_n \xrightarrow{P} X + Y$. 如果$g$连续，则$g(X_n) \xrightarrow{P} g(X)$，$g(X_n, Y_n) \xrightarrow{P} g(X, Y)$.

相合估计量的充分条件

定理：设总体$X \sim F(x;\theta)$，$\theta \in \Theta$，统计量$T(X_1,\dots,X_n)$为$g(\theta)$的估计量。如果

$$\lim_{n\to\infty} \mathbb{E}[T(X_1,\dots,X_n)] = g(\theta), \quad \lim_{n\to\infty} \mathrm{Var}[T(X_1,\dots,X_n)] = 0,$$

则$T(X_1,\dots,X_n)$为$g(\theta)$的相合估计量。

证明：令$T_n = T(X_1,\dots,X_n)$. 注意到

$$\{|T_n - g(\theta)| \geq \epsilon\} \subset \{|T_n - \mathbb{E}[T_n]| \geq \epsilon/2\} \cup \{|\mathbb{E}[T_n] - g(\theta)| \geq \epsilon/2\}.$$

对任意$\epsilon > 0$，存在N，当$n \geq N$时，$|\mathbb{E}[T_n] - g(\theta)| < \epsilon/2$. 此时

$$\{|T_n - g(\theta)| \geq \epsilon\} \subset \{|T_n - \mathbb{E}[T_n]| \geq \epsilon/2\}$$

所以，

$$P(|T_n - g(\theta)| \geq \epsilon) \leq P(|T_n - \mathbb{E}[T_n]| \geq \epsilon/2) \leq \frac{4\mathrm{Var}[T_n]}{\epsilon^2} \to 0.$$

目录

参数的区间估计

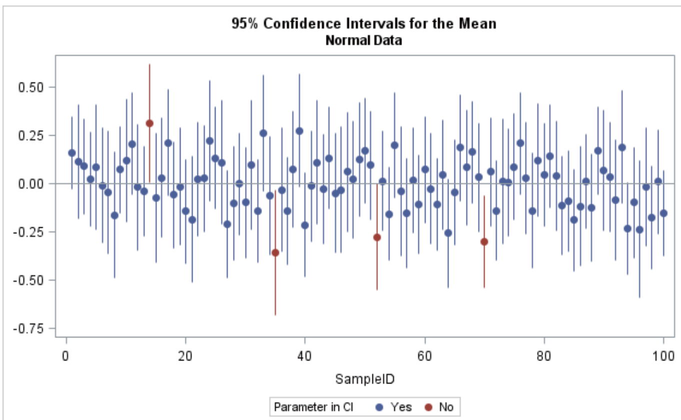
定义：设总体$X \sim F(x;\theta)$，$\theta \in \Theta$. 如果统计量$T_1(X_1,\dots,X_n)$，$T_2(X_1,\dots,X_n)$使得对给定的$\alpha \in (0,1)$有

$$P(T_1 \leq g(\theta) \leq T_2) = 1 - \alpha,$$

则称随机区间$[T_1, T_2]$为参数$g(\theta)$的置信度（置信概率）为$1-\alpha$的置信区间，T_1, T_2分别称为置信下界和置信上界。

注1：随机区间$[T_1, T_2]$包含$g(\theta)$的概率为$1-\alpha$. 在重复取样下，将得到许多不同的区间$[T_1, T_2]$，根据伯努利大数定理，这些区间中大约有$100(1-\alpha)\%$的区间包含未知参数。
注2：但对于一次抽样所得到的一个区间，决不能说"不等式$T_1 \leq g(\theta) \leq T_2$成立的概率为$1-\alpha$"。因为这时$T_1, T_2$是两个确定的数，从而只有两种可能，要么这个区间包含$g(\theta)$，要么这个区间不包含$g(\theta)$.

置信区间示意图



单个正态总体的期望的区间估计

(1) σ已知，求μ的置信区间

由抽样定理知，$\bar{X} \sim N(\mu, \sigma^2/n)$. 因此

$$U = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le u_{1-\alpha/2}\right) = 1-\alpha$$

即

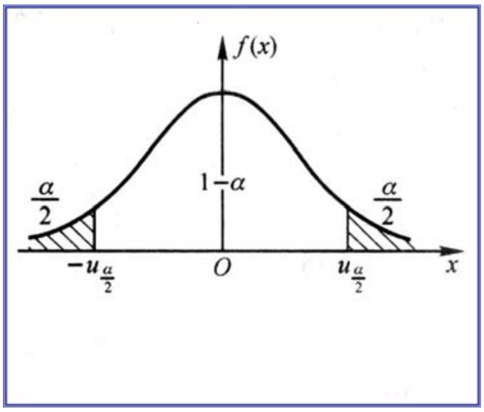
$$P\left(\bar{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

所以，区间

$$\left[\bar{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

为μ的置信度为$1-\alpha$的置信区间。

示意图



例

例：已知OPPO手机充电五分钟通话时间$X \sim N(\mu, 0.06)$. 随机抽取6部手机测试通话时间（单位：小时）为

1.6, 2.1, 1.9, 1.8, 2.2, 2.1,

求μ的置信度为95%的置信区间。

解：$\bar{X} = (1.6+\cdots+2.1)/6 = 1.95$. 已知$1-\alpha = 0.95$. 查表知，$u_{1-\alpha/2} = u_{0.975} = 1.96$, 置信区间为

$$\left[1.95 - 1.96\frac{\sqrt{0.06}}{\sqrt{6}},\ 1.95 + 1.96\frac{\sqrt{0.06}}{\sqrt{6}}\right] = [1.754,\ 2.146].$$

单个正态总体的期望的区间估计

(2) σ未知，求μ的置信区间

由抽样定理知，

$$T = \frac{\bar{X}-\mu}{S_n/\sqrt{n-1}} \sim t(n-1)$$

$$P\left(\left|\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}}\right| \le t_{1-\alpha/2}\right) = 1-\alpha$$

即

$$P\left(\bar{X} - t_{1-\alpha/2}\frac{S_n}{\sqrt{n-1}} \le \mu \le \bar{X} + t_{1-\alpha/2}\frac{S_n}{\sqrt{n-1}}\right) = 1-\alpha$$

所以，区间

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

为μ的置信度为$1-\alpha$的置信区间。

例

例：已知OPPO手机充电五分钟通话时间$X \sim N(\mu, \sigma^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

1.6, 2.1, 1.9, 1.8, 2.2, 2.1,

求μ的置信度为95%的置信区间。

解：$\bar{X} = (1.6+\cdots+2.1)/6 = 1.95$. $S_n = 0.206$. 已知$1-\alpha = 0.95$. 查表知，$t_{1-\alpha/2} = t_{0.975} = 2.5706$, 置信区间为

$$\left[1.95 - 2.5706\frac{0.206}{\sqrt{6-1}},\ 1.95 + 2.5706\frac{0.206}{\sqrt{6-1}}\right] = [1.713,\ 2.187].$$

一些思考...

- 分析这两种的结果会发现，由同一组样本观察值，按同样的置信概率，对μ计算出的置信区间因为σ的是否已知会不一样。这因为：当σ为已知时，我们掌握的信息多一些，在其他条件相同的情况下，对μ的估计精度要高一些，即表现为μ的置信区间长度要小些。反之，当σ为未知时，对μ的估计精度要低一些，即表现为μ的置信区间长度在大一些。
- 还可以发现，当样本量n不断增大时，两种情况下的置信区间会慢慢接近。也就意味着大样本信息可以弥补σ的缺失带来的偏差（大数定律）。

单个正态总体的方差的区间估计

(1) μ已知，求σ^2的置信区间

构造统计量

$$T=\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$$

$$P\left(\chi^2_{\alpha/2}(n)\le\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}\le\chi^2_{1-\alpha/2}(n)\right)=1-\alpha$$

即

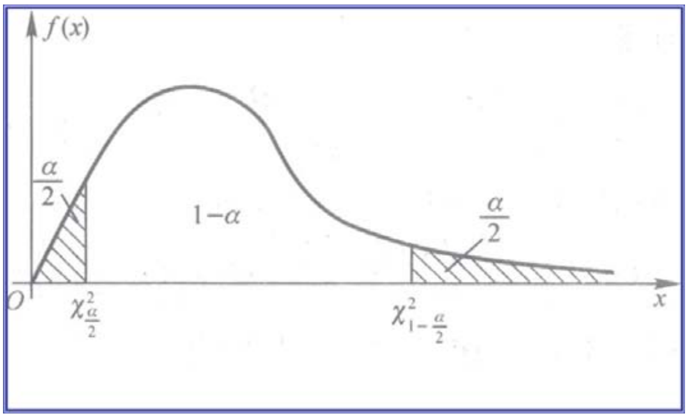
$$P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}\le\sigma^2\le\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}\right)=1-\alpha$$

所以，区间

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)},\ \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}\right]$$

为σ^2的置信度为$1-\alpha$的置信区间。

示意图



例

例： 已知OPPO手机充电五分钟通话时间$X\sim N(2,\sigma^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

1.6，2.1，1.9，1.8，2.2，2.1，

求σ^2的置信度为95%的置信区间。

解： $\sum_{i=1}^{n}(X_i-2)^2=0.27$. 已知$1-\alpha=0.95$. 查表知，$\chi^2_{\alpha/2}(6)=\chi^2_{0.025}(6)=1.24$, $\chi^2_{1-\alpha/2}(6)=\chi^2_{0.975}(6)=14.45$, 置信区间为

$$\left[\frac{0.27}{14.45},\ \frac{0.27}{1.24}\right]=[0.019,\ 0.218].$$

单个正态总体的方差的区间估计

(2) μ未知，求σ^2的置信区间

构造统计量

$$T=\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2}=\frac{nS_n^2}{\sigma^2}\sim\chi^2(n-1)$$

$$P\left(\chi^2_{\alpha/2}(n-1)\le\frac{nS_n^2}{\sigma^2}\le\chi^2_{1-\alpha/2}(n-1)\right)=1-\alpha$$

即

$$P\left(\frac{nS_n^2}{\chi^2_{1-\alpha/2}(n-1)}\le\sigma^2\le\frac{nS_n^2}{\chi^2_{\alpha/2}(n-1)}\right)=1-\alpha$$

所以，区间

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

为σ^2的置信度为$1-\alpha$的置信区间。

例

例： 已知OPPO手机充电五分钟通话时间$X\sim N(\mu,\sigma^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

1.6，2.1，1.9，1.8，2.2，2.1，

求σ^2的置信度为95%的置信区间。

解： $S_n^2=0.042$. 已知$1-\alpha=0.95$. 查表知，$\chi^2_{\alpha/2}(5)=\chi^2_{0.025}(5)=0.83$, $\chi^2_{1-\alpha/2}(5)=\chi^2_{0.975}(5)=12.83$, 置信区间为

$$\left[\frac{6\times0.042}{12.83},\ \frac{6\times0.042}{0.83}\right]=[0.020,\ 0.304].$$

两个正态总体期望之差的区间估计

设总体$X\sim N(\mu_1,\sigma_1^2)$, $(X_1,\ldots,X_m)$为其样本，样本均值为$\bar{X}$, 样本方差为S_{1m}^2; 另有与X独立的总体$Y\sim N(\mu_2,\sigma_2^2)$, $(Y_1,\ldots,Y_n)$为其样本，样本均值为$\bar{Y}$, 样本方差为S_{2n}^2.

(1) σ_1^2,σ_2^2已知，求$\mu_1-\mu_2$的置信区间

$$U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1).$$

$$P\left(\bar{X}-\bar{Y}-u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\le\mu_1-\mu_2\le\bar{X}-\bar{Y}+u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right)=1-\alpha$$

所以，区间

$$\left[(\bar{X}-\bar{Y})-u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}},\ (\bar{X}-\bar{Y})+u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right]$$

为$\mu_1-\mu_2$的置信度为$1-\alpha$的置信区间。

两个正态总体期望之差的区间估计

(2) σ_1^2, σ_2^2未知（已知$\sigma_1^2=\sigma_2^2=\sigma^2$），求$\mu_1-\mu_2$的置信区间

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/m+1/n}}\sim t(m+n-2).$$

$$P\left(\left|\bar{X}-\bar{Y}-(\mu_1-\mu_2)\right|\leq t_{1-\alpha/2}(m+n-2)S_w\sqrt{\frac{1}{m}+\frac{1}{n}}\right)=1-\alpha$$

令$t_{1-\alpha/2}(m+n-2)=t_{1-\alpha/2}$，区间

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

为$\mu_1-\mu_2$的置信度为$1-\alpha$的置信区间，其中

$$S_w=\sqrt{(mS_{1m}^2+nS_{2n}^2)/(m+n-2)}.$$

例

例： 已知OPPO手机充电五分钟通话时间$X\sim N(\mu_1,\sigma_1^2)$, VIVO手机充电五分钟通话时间$Y\sim N(\mu_2,\sigma_2^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

OPPO	1.6	2.1	1.9	1.8	2.2	2.1
VIVO	1.8	2.2	1.5	1.4	2.0	1.7

求$\mu_1-\mu_2$的置信度为95%的置信区间:

- 已知$\sigma_1^2=0.06$，$\sigma_2^2=0.08$.
- 已知$\sigma_1^2=\sigma_2^2$.

解： $m=n=6$, $\bar{X}=1.95$，$\bar{Y}=1.77$, $S_{1m}^2=0.042$, $S_{2n}^2=0.064$, $S_w=0.252$. 已知$1-\alpha=0.95$. 查表知，$u_{1-\alpha/2}=u_{0.975}=1.96$, $t_{1-\alpha/2}=t_{0.975}=2.5706$. 第一种情况置信区间为$[-0.12,\ 0.48]$. 第二种情况置信区间为$[-1.94,\ 0.554]$.

两个正态总体方差之比的区间估计

(1) μ_1, μ_2已知，求σ_1^2/σ_2^2的置信区间

$$T_1=\sum_{i=1}^{m}\frac{(X_i-\mu_1)^2}{\sigma_1^2}\sim\chi^2(m)$$

$$T_2=\sum_{i=1}^{n}\frac{(Y_i-\mu_2)^2}{\sigma_2^2}\sim\chi^2(n)$$

$$\frac{T_1/m}{T_2/n}=\frac{\frac{1}{m}\sum_{i=1}^m(X_i-\mu_1)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_2)^2}\frac{\sigma_2^2}{\sigma_1^2}\sim F(m,n)$$

$$P\left(F_{\alpha/2}(m,n)\leq\frac{\frac{1}{m}\sum_{i=1}^m(X_i-\mu_1)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_2)^2}\frac{\sigma_2^2}{\sigma_1^2}\leq F_{1-\alpha/2}(m,n)\right)=1-\alpha$$

所以，σ_1^2/σ_2^2的置信度为$1-\alpha$的置信区间为

$$\left[\frac{1}{F_{1-\alpha/2}(m,n)}\frac{\frac{1}{m}\sum_{i=1}^m(X_i-\mu_1)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_2)^2},\ \frac{1}{F_{\alpha/2}(m,n)}\frac{\frac{1}{m}\sum_{i=1}^m(X_i-\mu_1)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_2)^2}\right]$$

两个正态总体方差之比的区间估计

(2) μ_1, μ_2未知，求σ_1^2/σ_2^2的置信区间

$$T_1=\frac{(m-1)S_{1m}^{*2}}{\sigma_1^2}=\sum_{i=1}^{m}\frac{(X_i-\bar{X})^2}{\sigma_1^2}\sim\chi^2(m-1)$$

$$T_2=\frac{(n-1)S_{2n}^{*2}}{\sigma_2^2}=\sum_{i=1}^{n}\frac{(Y_i-\bar{Y})^2}{\sigma_2^2}\sim\chi^2(n-1)$$

$$\frac{T_1/(m-1)}{T_2/(n-1)}=\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\frac{\sigma_2^2}{\sigma_1^2}\sim F(m-1,n-1)$$

$$P\left(F_{\alpha/2}(m-1,n-1)\leq\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\frac{\sigma_2^2}{\sigma_1^2}\leq F_{1-\alpha/2}(m-1,n-1)\right)=1-\alpha$$

所以，σ_1^2/σ_2^2的置信度为$1-\alpha$的置信区间为

$$\left[\frac{1}{F_{1-\alpha/2}(m-1,n-1)}\frac{S_{1m}^{*2}}{S_{2n}^{*2}},\ \frac{1}{F_{\alpha/2}(m-1,n-1)}\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\right]$$

例

例： 已知OPPO手机充电五分钟通话时间$X\sim N(\mu_1,\sigma_1^2)$, VIVO手机充电五分钟通话时间$Y\sim N(\mu_2,\sigma_2^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

OPPO	1.6	2.1	1.9	1.8	2.2	2.1
VIVO	1.8	2.2	1.5	1.4	2.0	1.7

求σ_1^2/σ_2^2的置信度为90%的置信区间:

- 已知$\mu_1=2$，$\mu_2=1.8$.
- μ_1,μ_2未知.

解： $m=n=6$, $\sum_{i=1}^m(X_i-2)^2=0.27$，$\sum_{i=1}^n(Y_i-1.8)^2=0.46$, $S_{1m}^2=0.042$, $S_{2n}^2=0.064$. 已知$1-\alpha=0.9$. 查表知，$F_{1-\alpha/2}(6,6)=F_{0.95}(6,6)=4.28$, $F_{0.05}(6,6)=1/F_{0.95}(6,6)=0.23$. $F_{0.95}(5,5)=5.05$, $F_{0.05}(5,5)=1/F_{0.95}(5,5)=0.20$. 第一种情况置信区间为$[0.135,\ 2.512]$. 第二种情况置信区间为$[0.131,\ 3.314]$.

单个正态总体参数的联合区间估计

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|\leq\ell, k_1\leq\frac{nS_n^2}{\sigma^2}\leq k_2\right)=1-\alpha$$

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|\leq\ell\right)P\left(k_1\leq\frac{nS_n^2}{\sigma^2}\leq k_2\right)=1-\alpha$$

令

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|\leq\ell\right)=P\left(k_1\leq\frac{nS_n^2}{\sigma^2}\leq k_2\right)=\sqrt{1-\alpha}$$

取$\ell=u_{(1+\sqrt{1-\alpha})/2}$, $k_1=\chi^2_{(1-\sqrt{1-\alpha})/2}(n-1)$, $k_2=\chi^2_{(1+\sqrt{1-\alpha})/2}(n-1)$. 联合区间为

$$\left\{(\mu,\sigma^2):(\bar{X}-\mu)^2\leq\frac{\ell^2\sigma^2}{n},\ \frac{nSn^2}{k_2}\leq\sigma^2\leq\frac{nSn^2}{k_1}\right\}$$

非正态总体参数的区间估计

中心极限定理告诉我们

$$\frac{\bar{X}-\mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]/n}} \xrightarrow{d} N(0,1)$$

所以当n充分大的时候，

$$\frac{\bar{X}-\mathbb{E}[X]}{\sqrt{S_n^2/n}} = \frac{\sqrt{n}(\bar{X}-\mathbb{E}[X])}{S_n} \sim N(0,1)$$

所以期望$\mu=\mathbb{E}[X]$的置信度为$1-\alpha$的区间估计可以近似为

$$[\bar{X}-u_{1-\alpha/2}S_n/\sqrt{n}, \bar{X}+u_{1-\alpha/2}S_n/\sqrt{n}].$$

作业

第七章习题：6, 9, 13, 17, 18, 19, 20, 21

6月13日（周三）交！

概率论与数理统计

何志坚

华南理工大学数学学院

hezhijian@scut.edu.cn

第八章：假设检验

目录

女士品茶试验

奶茶是由牛奶与茶按一定比例混合而成，可以先倒茶后加奶，也可以先倒奶再倒茶。某女士声称她可以鉴别这两种混合方式，周围品茶的人对此产生了议论，都觉得不可思议。在场的费希尔也在思考这个问题，他提议做一项试验来检验如下命题是否可以接受：

假设H: 该女士无此种鉴别能力

他准备了10杯调好的奶茶（两种顺序的都有）给该女士鉴别，结果那位女士竟然能够正确地分辨出10杯奶茶中的每一杯的调制顺序。

如何做出你的判断？ 假如假设H是正确的，即该女士无此鉴别能力，她只能靠猜，每次猜对的概率是$1/2$，连续10次猜对的概率为$2^{-10}<0.001$, 这是一个很小的概率，在一次试验中几乎不会发生，如今发生了，只能说明假设H不成立。假如该女士只猜对了9杯（或者8杯），又该如何判断?

参数假设检验基本概念

设有来自某一参数分布族$\{F(x,\theta)\}$, $\theta\in\Theta$, 其中Θ为参数空间。
原假设（零假设）：

$$H_0:\theta\in\Theta_0$$

备选假设（对立假设，备择假设）：

$$H_1:\theta\in\Theta_1$$

其中$\varnothing\neq\Theta_0,\Theta_1\subset\Theta$，$\Theta_0\cap\Theta_1=\varnothing$. 最常见的情况$\Theta_1=\Theta-\Theta_0$.

简单原假设：Θ_0只包含一个点，如$H_0:\theta=\theta_0$, 备选假设H_1通常有三种可能：

- 双边假设：$H_1:\theta\neq\theta_0$
- 单边假设：$H_1:\theta>\theta_0$
- 单边假设：$H_1:\theta<\theta_0$

其它情况称为**复杂（复合）原假设**。

假设检验

设X为通话时间，$X\sim N(\mu,\sigma^2)$.

$$H_0:\mu=2\ vs.\ H_1:\mu\neq 2$$

H_0为原假设（简单原假设）
H_1为备选假设
该假设为双边假设，如何检验?

假设检验基本原则

检验法则：检验本质上是把样本空间划分成两个互不相交的部分W和$\bar{W}$, 当样本属于W时就拒绝H_0; 否则接受H_0. 称W为该检验的拒绝域，而$\bar{W}$为接受域。

例：假设检验$H_0:\mu=2\ vs.\ H_1:\mu\neq 2$, 拒绝域可以设为

$$W=\{(x_1,\ldots,x_n):|\bar{x}-2|\geq c\}.$$

c称为临界值。

小概率原理：小概率事件在一次试验中是几乎不发生的。换言之，如果H_0为真的前提下，样本落在拒绝域是一个小概率事件，不应该发生。如果发生，则推翻原命题，即拒绝原假设。

给定一个比较小的数$\alpha\in(0,1)$, 通常$\alpha=\{0.01,0.05,0.001\}$, 可求$c$使得

$$P((X_1,\ldots,X_n)\in W|H_0\text{为真})=\alpha$$

称值α为显著性水平(或检验水平)，用来衡量原假设与实际情况差异是否明显的标准。

显著性水平的选择

人们自然会产生这样的问题：概率小到什么程度才当作“小概率事件”呢？这要据实际情况而定，例如即使下雨的概率为10%，仍有人会因为它太小而不带雨具。但某航空公司的事故率为1%，人们就会因为它太大而不敢乘坐该公司的飞机，通常把概率不超过0.05 (或0.01)的事件当作“小概率事件”。为此在假设检验时，必须先确定小概率即显著性的值α (即不超过α的概率认为是小概率)。

两类错误

第一类错误：H_0正确，但拒绝了它，这类错误也称“拒真错误”。

$$\alpha=P(\text{拒绝}H_0|H_0\text{为真})=P((X_1,\ldots,X_n)\in W|H_0\text{为真})$$

第二类错误：H_0不正确(即H_1为真)，但接受了它，这类错误也称“受伪错误”。

$$\beta=P(\text{接受}H_0|H_1\text{为真})=P((X_1,\ldots,X_n)\in\bar{W}|H_1\text{为真})$$

- $\beta\neq 1-\alpha$, 可以证明，在样本容量一定时，不可能同时缩小两类错误。
- 当样本容量一定时，犯第一类错误的概率越小，则犯第二类错误的概率越大。
- 当现实中样本容量不可能无限制的大，从而同时控制两类错误就不可能。

显著性检验

实际中常用的是只控制第一类错误而不控制第二类错误的检验方法，即显著性检验。当想用显著性检验对某一猜测结论作强有力的支持时，应该将猜测结论的反面作为原假设。这是反证法的思想。

例

例：设总体$X \sim N(\mu, \sigma^2)$，σ已知，$(X_1, \dots, X_n)$为其样本，对假设检验问题

$$H_0 : \mu = \mu_0,\ H_1 : \mu = \mu_1$$

做检验，其中$\mu_0 < \mu_1$，解释该检验的第一类错误的概率α与第二类错误的概率β之间的关系。

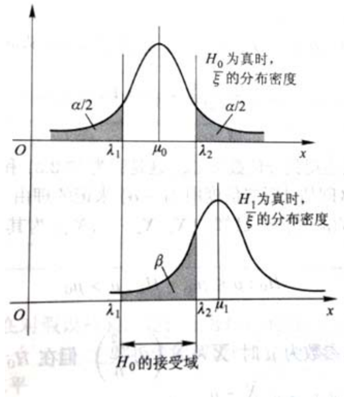
解：选择拒绝域为$W = \{|\bar{X} - \mu_0| \geq c\}$. 则有

$$\alpha = P(|\bar{X} - \mu_0| \geq c | \mu = \mu_0) = P\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \geq u_{1-\alpha/2}\right)$$

接受域$\bar{W} = \{\mu_0 - u_{1-\alpha/2}\sigma/\sqrt{n} < \bar{X} < \mu_0 + u_{1-\alpha/2}\sigma/\sqrt{n}\} =: \{\lambda_1 < \bar{X} < \lambda_2\}$.

$$\beta = P(\bar{X} \in \bar{W} | \mu = \mu_1) = \Phi((\lambda_2 - \mu_1)\sqrt{n}/\sigma) - \Phi((\lambda_1 - \mu_1)\sqrt{n}/\sigma).$$

两类错误关系的图示



假设检验的基本步骤

1. 提出假设
2. 选择检验统计量，给出拒绝域形式
3. 求临界值（确定接受域/拒绝域）
4. 算出观察值
5. 作出判断

。

目录

单个正态总体期望的假设检验（双边）

(1)已知方差σ^2

$$H_0 : \mu = \mu_0,\ H_1 : \mu \neq \mu_0$$

选择检验统计量：

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

求临界值：

$$P(|U| \geq u_{1-\alpha/2}) = \alpha.$$

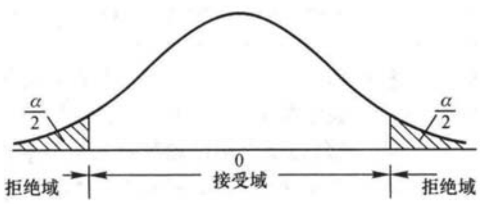
接受域：

$$\bar{W} = \{|U| < u_{1-\alpha/2}\} = \{\mu_0 - u_{1-\alpha/2}\sigma/\sqrt{n} < \bar{X} < \mu_0 + u_{1-\alpha/2}\sigma/\sqrt{n}\}$$

算出观察值：根据样本的观测值算出统计量U的观测值u

作出判断：若$u < u_{1-\alpha/2}$，则接受原假设，若$u \geq u_{1-\alpha/2}$，则拒绝原假设，接受备选假设。这种检验方法称为u检验法。

图例



注：α越大越容易拒绝原假设。

例

例：已知OPPO手机充电五分钟通话时间$X \sim N(\mu, 0.07^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

$$1.6,\ 2.1,\ 1.9,\ 1.8,\ 2.2,\ 2.1,$$

检验OPPO手机的平均通话时间为2小时是否成立，显著性水平取为5%.

解：
(1)提出假设：$H_0: \mu = 2\ vs.\ H_1: \mu \neq 2$
(2)选择检验统计量：$U = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$
(3)求临界值：$P(|U| \geq u_{1-\alpha/2}) = \alpha = 0.05$, $u_{1-\alpha/2} = u_{0.975} = 1.96$.
(4)算出观察值：$\bar{x} = (1.6 + \cdots + 2.1)/6 = 1.95$, $u = \frac{|1.95-2|}{0.07/\sqrt{6}} = 1.75 < 1.96$
(5)作出判断：接受原假设
思考：如果$\alpha = 10\%$呢？$u_{0.95} = 1.65$ 拒绝原假设

检验的p值

从上个例子中发现，对于同一个的样本观测值，在一个较小($\alpha = 5\%$)的显著水平下得到接受原假设的结论，而在一个较大($\alpha = 10\%$)的显著水平下得到拒绝原假设的结论。
问题：是否存在一个数$p \in (0,1)$，使得

- 当$\alpha \geq p$时，拒绝H_0
- 当$\alpha < p$时，接受H_0

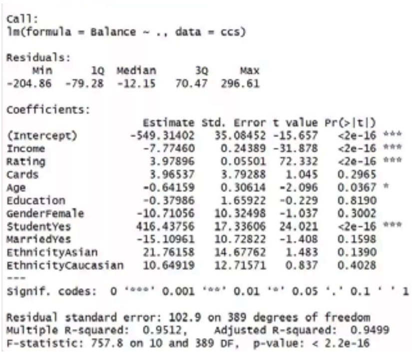
定义：在一个假设检验问题中，利用样本观测值能够作出拒绝原假设的最小显著水平称为检验的p值。

对于上例，$p = \min\{\alpha : u_{1-\alpha/2} \leq |u|\}$，即$u_{1-p/2} = |u| = 1.75$，于是

$$p = P(|U| \geq 1.75) = 2(1 - \Phi(1.75)) = 8\%$$

注：统计软件一般只给出p值，而不是针对给定的显著性水平α进行判断。

R软件中p值的解释



```
Call:
lm(formula = Balance ~ ., data = ccs)

Residuals:
    Min      1Q  Median      3Q     Max
-204.86  -79.28  -12.15   70.47  296.61

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -549.31402   35.08452 -15.657   <2e-16 ***
Income              -7.77460    0.24389 -31.878   <2e-16 ***
Rating               3.97896    0.05501  72.332   <2e-16 ***
Cards                3.96537    3.79288   1.045   0.2965
Age                 -0.64159    0.30614  -2.096   0.0367 *
Education           -0.37986    1.65922  -0.229   0.8190
GenderFemale       -10.71056   10.32498  -1.037   0.3002
StudentYes         416.43756   17.33606  24.021   <2e-16 ***
MarriedYes         -15.10961   10.72822  -1.408   0.1598
EthnicityAsian      21.76158   14.67762   1.483   0.1390
EthnicityCaucasian  10.64919   12.71571   0.837   0.4028
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 102.9 on 389 degrees of freedom
Multiple R-squared:  0.9512,    Adjusted R-squared:  0.9499
F-statistic: 757.8 on 10 and 389 DF,  p-value: < 2.2e-16
```


单个正态总体期望的假设检验（双边）

(2)未知方差σ^2

$$H_0: \mu = \mu_0,\ H_1: \mu \neq \mu_0$$

选择检验统计量：

$$T = \frac{\bar{X} - \mu_0}{S_n/\sqrt{n-1}} \sim t(n-1)$$

求临界值：

$$P(|T| \geq t_{1-\alpha/2}) = \alpha.$$

接受域：

$$\bar{W} = \{|T| < t_{1-\alpha/2}(n-1)\}$$

算出观察值：根据样本的观测值算出统计量T的观测值t
作出判断：若$t < t_{1-\alpha/2}(n-1)$，则接受原假设，若$t \geq t_{1-\alpha/2}(n-1)$，则拒绝原假设，接受备选假设。这种检验方法称为t检验法。

例

例：已知OPPO手机充电五分钟通话时间$X \sim N(\mu, \sigma^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

$$1.6,\ 2.1,\ 1.9,\ 1.8,\ 2.2,\ 2.1,$$

检验OPPO手机的平均通话时间为2小时是否成立，显著性水平取为5%.

解：
(1)提出假设：$H_0: \mu = 2\ vs.\ H_1: \mu \neq 2$
(2)找检验统计量：$\frac{\bar{X}-2}{S_n/\sqrt{5}} \sim t(5)$
(3)求临界值：$P(|T| \geq t_{1-\alpha/2})(5) = \alpha = 0.05$, $t_{0.975}(5) = 2.5706$.
(4)算出观察值：$\bar{x} = 1.95$, $s_n = 0.206$, $T = \left|\frac{1.95-2}{0.206/\sqrt{5}}\right| = 0.54 < 2.5706$
(5)作出判断：接受原假设

单个正态总体期望的假设检验对比

序号	H_0	H_1	σ^2已知	σ^2未知
I	$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\lvert\bar{x}-\mu_0\rvert}{\sigma/\sqrt{n}} \geq u_{1-\alpha/2}$	$\frac{\lvert\bar{x}-\mu_0\rvert}{S_n/\sqrt{n-1}} \geq t_{1-\alpha/2}(n-1)$
II	$\mu = \mu_0$	$\mu > \mu_0$	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geq u_{1-\alpha}$	$\frac{\bar{X}-\mu_0}{S_n/\sqrt{n-1}} \geq t_{1-\alpha}(n-1)$
III	$\mu \leq \mu_0$	$\mu > \mu_0$		
IV	$\mu = \mu_0$	$\mu < \mu_0$	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leq u_{\alpha}$	$\frac{\bar{X}-\mu_0}{S_n/\sqrt{n-1}} \leq t_{\alpha}(n-1)$
V	$\mu \geq \mu_0$	$\mu < \mu_0$		

Table: 单个正态总体均值μ的假设检验的拒绝域（显著性水平为α）

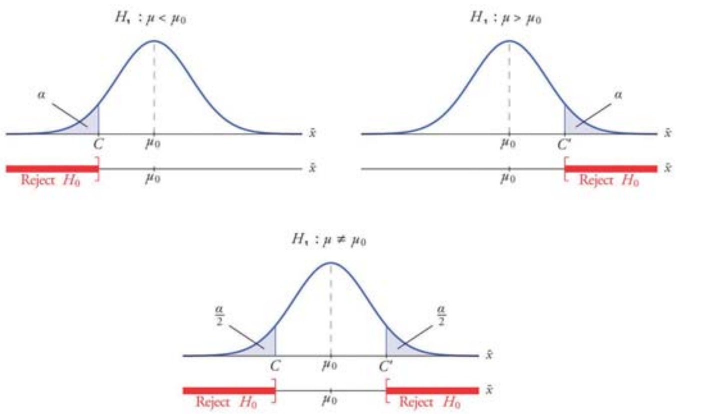
注：对于情形III, 犯第一类错误的概率为

$$P(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geq u_{1-\alpha}|\mu \leq \mu_0) \leq P(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geq u_{1-\alpha}|\mu \leq \mu_0) = \alpha$$

同理，对于情形V, 犯第一类错误的概率同样不超过α.

思考：各种情况下的p值如何计算？

单边假设与双边假设



假设检验与置信区间的关系

考虑双侧检验(σ^2已知)，显著性水平为α的接受域为

$$\bar{W} = \{\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \leq u_{1-\alpha/2}\} = \{\bar{X} - u_{1-\alpha/2}\sigma/\sqrt{n} \leq \mu_0 \leq \bar{X} + u_{1-\alpha/2}\sigma/\sqrt{n}\}$$

为μ的$1-\alpha$置信区间。这表明，假设检验与置信区间是一一对应的。

单个正态总体方差的假设检验（双边）

(1)已知期望μ

$$H_0 : \sigma^2 = \sigma_0^2,\ H_1 : \sigma^2 \neq \sigma_0^2$$

选择检验统计量：

$$\chi^2 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$$

求临界值：

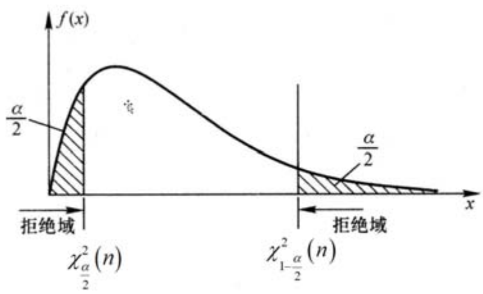
$$P(\chi^2 \geq \chi^2_{1-\alpha/2}(n)) = P(\chi^2 \leq \chi^2_{\alpha/2}(n)) = \alpha/2.$$

接受域：

$$\bar{W} = \{\chi^2_{\alpha/2}(n) < \chi^2 < \chi^2_{1-\alpha/2}(n)\}$$

算出观察值：根据样本的观测值算出统计量χ^2的观测值χ_1^2
作出判断：若$\chi^2_{\alpha/2}(n) < \chi_1^2 < \chi^2_{1-\alpha/2}(n)$, 则接受原假设，否则拒绝原假设，接受备选假设。这种检验方法称为χ^2检验法。

图例



单个正态总体方差的假设检验（双边）

(2)未知期望μ

$$H_0 : \sigma^2 = \sigma_0^2,\ H_1 : \sigma^2 \neq \sigma_0^2$$

选择检验统计量：

$$\chi^2 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$$

求临界值：

$$P(\chi^2 \geq \chi^2_{1-\alpha/2}(n-1)) = P(\chi^2 \leq \chi^2_{\alpha/2}(n-1)) = \alpha/2.$$

接受域：

$$\bar{W} = \{\chi^2_{\alpha/2}(n-1) < \chi^2 < \chi^2_{1-\alpha/2}(n-1)\}$$

算出观察值：根据样本的观测值算出统计量χ^2的观测值χ_1^2
作出判断：若$\chi^2_{\alpha/2}(n-1) < \chi_1^2 < \chi^2_{1-\alpha/2}(n-1)$, 则接受原假设，否则拒绝原假设，接受备选假设。

例

例：已知OPPO手机充电五分钟通话时间$X \sim N(\mu, \sigma^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

1.6, 2.1, 1.9, 1.8, 2.2, 2.1,

检验OPPO手机的通话时间方差为0.06是否成立，显著性水平取为5%:

- 已知$\mu = 2$.
- 未知μ.

解：假设检验：

$$H_0 : \sigma^2 = 0.06,\ H_1 : \sigma^2 \neq 0.06$$

(1)检验统计量的观测值为$\frac{1}{0.06}\sum_{i=1}^{n}(x_i - 2)^2 = 4.5 \in (1.24, 14.45)$. 接受$H_0$
(2)检验统计量的观测值为$\frac{1}{0.06}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 4.2 \in (0.83, 12.83)$. 接受$H_0$

单个正态总体方差的假设检验对比

序号	H_0	H_1	μ已知	μ未知
I	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\le\chi^2_{\frac{\alpha}{2}}(n)$ 或 $\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\ge\chi^2_{1-\frac{\alpha}{2}}(n)$	$\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\le\chi^2_{\frac{\alpha}{2}}(n-1)$或 $\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\ge\chi^2_{1-\frac{\alpha}{2}}(n-1)$
II	$\sigma^2=\sigma_0^2$	$\sigma^2>\sigma_0^2$	$\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\ge\chi^2_{1-\alpha}(n)$	$\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\ge\chi^2_{1-\alpha}(n-1)$
III	$\sigma^2\le\sigma_0^2$	$\sigma^2>\sigma_0^2$		
IV	$\sigma^2=\sigma_0^2$	$\sigma^2<\sigma_0^2$	$\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\le\chi^2_{\alpha}(n)$	$\frac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma_0^2}\le\chi^2_{\alpha}(n-1)$
V	$\sigma^2\ge\sigma_0^2$	$\sigma^2<\sigma_0^2$		

Table: 单个正态总体均值σ^2的假设检验的拒绝域（显著性水平为α）

两个独立正态总体均值差的假设检验（双边）

设总体$X\sim N(\mu_1,\sigma_1^2)$, 另有与$X$独立的总体$Y\sim N(\mu_2,\sigma_2^2)$.

$$H_0:\mu_1-\mu_2=\delta,\ H_1:\mu_1-\mu_2\neq\delta$$

(1) σ_1^2,σ_2^2已知

选择检验统计量:

$$U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}=\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1)$$

若$|U|\ge u_{1-\alpha/2}$拒绝H_0，否则接受H_0.

(2) σ_1^2,σ_2^2未知，已知$\sigma_1^2=\sigma_2^2$

选择检验统计量:

$$T=\frac{\bar{X}-\bar{Y}-\delta}{S_w\sqrt{1/m+1/n}}\sim t(m+n-2)$$

若$|T|\ge t_{1-\alpha/2}(n+m-2)$拒绝$H_0$，否则接受$H_0$.

例

例: 已知OPPO手机充电五分钟通话时间$X\sim N(\mu_1,\sigma_1^2)$, VIVO手机充电五分钟通话时间$Y\sim N(\mu_2,\sigma_2^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

OPPO	1.6	2.1	1.9	1.8	2.2	2.1
VIVO	1.8	2.2	1.5	1.4	2.0	1.7

问这两种手机通话时间平均水平无显著差异（显著性水平$\alpha=0.05$）：

- 已知$\sigma_1^2=0.06$, $\sigma_2^2=0.08$.
- 已知$\sigma_1^2=\sigma_2^2$.

解: 假设检验:

$$H_0:\mu_1=\mu_2,\ H_1:\mu_1\neq\mu_2$$

(1)检验统计量的观测值为$\frac{\bar{x}-\bar{y}}{\sqrt{0.06/6+0.08/6}}=1.18<u_{0.975}=1.96$. 接受$H_0$

(2)检验统计量的观测值为$\frac{\bar{x}-\bar{y}}{s_w\sqrt{1/6+1/6}}=1.24<t_{0.975}(10)=2.23$. 接受$H_0$

两个独立正态总体均值差的假设检验对比

序号	H_0	H_1	σ_1^2,σ_2^2已知	σ_1^2,σ_2^2未知
I	$\mu_1-\mu_2=\delta$	$\mu_1-\mu_2\neq\delta$	$\frac{\lvert\bar{x}-\bar{y}-\delta\rvert}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\ge u_{1-\frac{\alpha}{2}}$	$\frac{\lvert\bar{x}-\bar{y}-\delta\rvert}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\ge t_{1-\frac{\alpha}{2}}$
II	$\mu_1-\mu_2=\delta$	$\mu_1-\mu_2>\delta$	$\frac{\bar{x}-\bar{y}-\delta}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\ge u_{1-\frac{\alpha}{2}}$	$\frac{\bar{x}-\bar{y}-\delta}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\ge t_{1-\alpha}$
III	$\mu_1-\mu_2\le\delta$	$\mu_1-\mu_2>\delta$		
IV	$\mu_1-\mu_2=\delta$	$\mu_1-\mu_2<\delta$	$\frac{\bar{x}-\bar{y}-\delta}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\le u_{\alpha}$	$\frac{\bar{x}-\bar{y}-\delta}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\le t_{\alpha}$
V	$\mu_1-\mu_2\ge\delta$	$\mu_1-\mu_2<\delta$		

Table: 单个正态总体均值μ的假设检验的拒绝域（显著性水平为α），这里t分布的分位数对应的自由度为$m+n-2$. $s_w=\sqrt{(ms_{1m}^2+ns_{2n}^2)/(n+m-2)}$

两个独立正态总体方差的假设检验（双边）

设总体$X\sim N(\mu_1,\sigma_1^2)$, 另有与$X$独立的总体$Y\sim N(\mu_2,\sigma_2^2)$.

$$H_0:\sigma_1^2=\sigma_2^2,\ H_1:\sigma_1^2\neq\sigma_2^2$$

(1) μ_1,μ_2已知

选择检验统计量:

$$F=\frac{\frac{1}{m}\sum_{i=1}^m(X_i-\mu_1)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_2)^2}\frac{\sigma_2^2}{\sigma_1^2}=\frac{\frac{1}{m}\sum_{i=1}^m(X_i-\mu_1)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_2)^2}\sim F(m,n)$$

若$F\ge F_{1-\alpha/2}(m,n)$或者$F<F_{\alpha/2}(m,n)$拒绝H_0，否则接受H_0.

(2) μ_1,μ_2未知

选择检验统计量:

$$F=\frac{\frac{1}{m-1}\sum_{i=1}^m(X_i-\bar{X})^2}{\frac{1}{n-1}\sum_{i=1}^n(Y_i-\bar{Y})^2}=\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\sim F(m-1,n-1)$$

若$F\ge F_{1-\alpha/2}(m-1,n-1)$或者$F<F_{\alpha/2}(m-1,n-1)$拒绝$H_0$，否则接受$H_0$.

这种检验称为F检验法。

例

例: 已知OPPO手机充电五分钟通话时间$X\sim N(\mu_1,\sigma_1^2)$, VIVO手机充电五分钟通话时间$Y\sim N(\mu_2,\sigma_2^2)$. 随机抽取6部手机测试通话时间（单位：小时）为

OPPO	1.6	2.1	1.9	1.8	2.2	2.1
VIVO	1.8	2.2	1.5	1.4	2.0	1.7

求检验两种手机通话时间方差是否相等（显著性水平$\alpha=0.1$）：

- 已知$\mu_1=2$, $\mu_2=1.8$.
- μ_1,μ_2未知.

解: 假设检验:

$$H_0:\sigma_1^2=\sigma_2^2,\ H_1:\sigma_1^2\neq\sigma_2^2$$

(1)检验统计量的观测值为$\frac{\frac{1}{m}\sum_{i=1}^m(X_i-2)^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-1.8)^2}=0.59\in(0.23,4.28)$. 接受$H_0$

(2)检验统计量的观测值为$\frac{\frac{1}{m}\sum_{i=1}^m(X_i-\bar{X})^2}{\frac{1}{n}\sum_{i=1}^n(Y_i-\bar{Y})^2}=0.43\in(0.2,5.05)$. 接受$H_0$

两个独立正态总体方差的假设检验对比

序号	H_0	H_1	μ已知	μ未知
I	$\sigma_1^2=\sigma_2^2$	$\sigma_1^2\neq\sigma_2^2$	$\frac{\sum_{i=1}^m (x_i-\mu_1)^2/m}{\sum_{i=1}^n (y_i-\mu_2)^2/n}\leq F_{\frac{\alpha}{2}}$ 或 $\frac{\sum_{i=1}^m (x_i-\mu_1)^2/m}{\sum_{i=1}^n (y_i-\mu_2)^2/n}\geq F_{1-\frac{\alpha}{2}}$	$\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\leq F_{\frac{\alpha}{2}}$ 或 $\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\geq F_{1-\frac{\alpha}{2}}$
II	$\sigma_1^2=\sigma_2^2$	$\sigma_1^2>\sigma_2^2$	$\frac{\sum_{i=1}^m (x_i-\mu_1)^2/m}{\sum_{i=1}^n (y_i-\mu_2)^2/n}\geq F_{1-\alpha}$	$\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\geq F_{1-\alpha}$
III	$\sigma_1^2\leq\sigma_2^2$	$\sigma_1^2>\sigma_2^2$		
IV	$\sigma_1^2=\sigma_2^2$	$\sigma_1^2<\sigma_2^2$	$\frac{\sum_{i=1}^m (x_i-\mu_1)^2/m}{\sum_{i=1}^n (y_i-\mu_2)^2/n}\leq F_{\alpha}$	$\frac{S_{1m}^{*2}}{S_{2n}^{*2}}\leq F_{\alpha}$
V	$\sigma_1^2\geq\sigma_2^2$	$\sigma_1^2<\sigma_2^2$		

Table: 两个独立正态总体方差假设检验的拒绝域（显著性水平为α），倒数第二列F分布为$F(m,n)$，最后一列F分布为$F(m-1,n-1)$

目录

非正态总体均值的假设检验

与非正态总体的区间估计思想类型，当样本量充分大时，以下检验统计量可近似看出标准正态分布：

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1).$$

如果σ未知，可以用样本标准差S_n替代σ.

非参假设检验

前面讨论的关于参数的假设检验，都是事先假定总体的分布类型为已知的。但有些时候，事先并不知道总体服从什么分布，需要对总体的分布类型进行推断。这类检验称为非参数假设检验。

$$H_0: X\sim F_0(\cdot;\boldsymbol{\theta}),\ \ H_1: X\nsim F_0(\cdot;\boldsymbol{\theta})$$

例：比如检验某总体是否为正态总体

常用的方法：χ^2拟合优度检验——构造皮尔逊(Pearson)统计量

自学

最后一次作业

根据期中考试成绩数据回答调查问卷中的问题，6月22日（周五）前提交！

数据链接：`https://www.jianguoyun.com/p/DenoTWkQpvLJBhj-y1o`
问卷连接：`https://ks.wjx.top/jq/24951942.aspx`