

某不科学的概统复习提纲^{*}

Written by 一个不靠谱的人

他不是

Michael Zhu

才怪[†]

1st version was completed on 2015.12.29

Last edited on 2016.4.22[‡]

^{*}如发现错误，请联系作者 [idleleavesleaf\(at\)gmail.com](mailto:idleleavesleaf(at)gmail.com)

[†]这是四个人，不是一个人，请不要理解错。

[‡]Recompiled using X₃T₃X with new fonts. All the texts modified are marked red.

最高指示

我的经历就是到了上海，到了89年年初的时候，我在想我估计也快要离休了，我想我应该去当教授。于是我就给朱物华校长、张钟俊院长，给他们写了一个报告。他们说欢迎你来，不过，这个Apply for Professor，你要去做一个报告。我做了一个能源与发展趋势的主要的节能措施，这个报告经过好几百个教授一致通过。那么上海交大教授当了以后我就做第二个报告，微电子工业的发展。这两个报告做了以后不久，过后，1989年的5月31号北京就把我调到北京去了。现在这个报告做了也快20年了，所以去年我就在我们交大的学报发表了两篇文章，就是呼应89年的报告的。特别是昨天晚上，他们又把我第二篇报告，还有我这十几年包括在电子工业部、上海市所做的有关于信息产业化的文章，总共我听他们讲是27篇。我也没有什么东西可以送给你们的，我拿到以后我叫钱秘书啊，就把这两个学报，两个学报的英文本，因为他们这里洋文好的人多得很呐，还有前面出过两本书，加上昨天晚上出的这本书，送给郭伟华同志，给你送过来，给你们做个纪念。

人呐就都不知道，自己就不可以预料。一个人的命运啊，当然要靠自己奋斗，但是也要考虑到历史的行程，我绝对不知道，我作为一个上海市委书记怎么把我选到北京去了，所以邓小平同志同我讲话，说“中央都决定了，你来当总书记”，我说另请高明吧。我实在我也不是谦虚，我一个上海市委书记怎么到北京来了呢？但是呢，小平同志讲“大家已经研究决定了”，所以后来我就念了两首诗，叫“苟利国家生死以，岂因祸福避趋之”，所以我就到了北京。

到了北京我干了这十几年也没有什么别的，大概三件事：

一个，确立了社会主义市场经济；

第二个，把邓小平理论列入党章；

第三个，就是“三个代表”。

如果说还有一点成绩就是军队一律不得经商，这个对军队的命运有很大的关系。因为我后来又干了一年零八个月，等于我在部队干了15年军委主席。还有九八年的抗洪也是很大的。但这些都是次要的，我主要的我就是三件事情，很惭愧，就做了一点微小的工作，谢谢大家。

Contents

1 事件的概率	4
1.1 概率是什么	4
1.2 古典概率计算	5
1.3 事件的运算、条件概率与独立性	5
2 随机变量及概率分布	8
2.1 一维随机变量	8
2.1.1 离散型	8
2.1.2 连续型	9
2.2 多维随机变量 (随机向量)	11
2.2.1 离散型	11
2.2.2 连续型	12
2.2.3 边缘分布	13
2.3 条件概率分布与随机变量的独立性	13
2.3.1 离散型	13
2.3.2 连续型	14
2.3.3 随机变量的独立性	14
2.4 随机变量的函数的概率分布	15
2.4.1 离散型	15
2.4.2 连续型 (一般讨论)	15
2.4.3 随机变量和的密度函数	16
2.4.4 随机变量商的密度函数	17
3 随机变量的数字特征	19
3.1 数学期望 (均值) 与中位数	19
3.1.1 数学期望	19
3.2 方差与矩	21
3.3 协方差与相关系数	22
3.4 大数定理和中心极限定理	23

4	参数估计	25
4.1	数理统计学的基本概念	25
4.2	矩估计、极大似然估计	25
4.2.1	矩估计法	25
4.2.2	极大似然估计法	26
4.3	点估计的优良性准则	26
4.3.1	最小方差无偏估计	27
4.4	区间估计	27
4.4.1	枢轴变量法	27
4.4.2	大样本法	29
4.4.3	置信界	29
5	假设检验	31
5.1	问题提法和基本概念	31
5.2	重要参数检验	33
5.2.1	一样本正态总体 $N(\mu, \sigma^2)$ 的假设检验	33
5.2.2	两样本正态总体的假设检验	34
5.2.3	成对数据	34
5.2.4	二项分布中未知参数 p 的假设检验	35
5.3	拟合优度检验	35
5.3.1	离散分布且理论分布完全已知	35
5.3.2	离散分布且理论分布含若干未知参数	36
5.3.3	离散分布列联表的独立性检验	36
5.3.4	离散分布列联表的齐一性检验	36
5.3.5	连续总体情形	36
6	Acknowledgement	37
7	References	37

1 事件的概率

1.1 概率是什么

概率 某种情况（事件）出现的可能性大小的一种数量指标，介于0到1之间。^{*†}

事件

(1) 有一个明确界定的试验。

(2) 这个试验的全部可能结果，是在试验前就明确的。

(3) 我们有一个明确的陈述，这个陈述界定了试验的全部可能结果中一个确定的部分。这个陈述，或者说一个确定的部分，就叫做一个事件。

概率（古典概率） 设一个试验有 N 个等可能的结果，而事件 E 恰包含其中的 M 个结果，则事件 E 的概率，记为 $P(E)$ ，定义为：

$$P(E) = M/N$$

推广：几何概率 基于几何图形的面积、体积、长度等而算出的。

统计意义 频率（大量次数上重复施行）逼近概率

概率的公理化定义

(1) $0 \leq P(A) \leq 1$.

(2) $P(\Omega) = 1, P(\emptyset) = 0$.

^{*}The 1st version of this chapter was completed on 2015.12.22

[†]Last edited on 2016.4.22

(3) 加法公理

1.2 古典概率计算

排列组合的几个简单公式

$$(1) P_r^n = \frac{n!}{(n-r)!}, \text{ 特别的, } P_r^r = r!.$$

$$(2) C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

$$(3) (a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}.$$

(4) n 个相异物件分成 k 堆, 各堆物件数分别为 r_1, \dots, r_k 的分法是:

$$n! / r_1! \cdots r_k!$$

1.3 事件的运算、条件概率与独立性

包含 $A \subset B$ ($B \supset A$) 同集合论内容。

相等 $A = B$ 同集合论内容。

对立事件 $B = \{A \text{ 不发生} \} = \bar{A}$

和 (并) $C = \{A \text{ 发生, 或 } B \text{ 发生} \} = A + B$

推广 $C = \sum_{i=1}^n A_i$

概率的加法定理 若干个互斥事件之和的概率, 等于各事件的概率之和, 即

$$P(A_1 + A_2 + \cdots) = P(A_1) + P(A_2) + \cdots.$$

推论 $P(\bar{A}) = 1 - P(A)$.

积 (交) $C = \{A, B \text{ 都发生}\} = AB$

推广 $C = \prod_{i=1}^n A_i$

差 $C = \{A \text{ 发生, } B \text{ 不发生}\} = A - B$

和、积、差的一些结论

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

(2) 和、积的交换律, 结合律。

(3) 分配律, 例: $A(B - C) = AB - AC$.

(4) $\overline{\prod_{i=1}^n A_i} = \sum_{i=1}^n \bar{A}_i$, $\overline{\sum_{i=1}^n A_i} = \prod_{i=1}^n \bar{A}_i$

注意: $A + A = A \neq 2A$, $AA = A$, $(A - B) + B = A + B \neq A$.

本质上是一种逻辑关系!

条件概率 设有两个事件 A, B , 而 $P(B) \neq 0$. 则“在给定 B 发生的情况下 A 的条件概率”, 记为 $P(A|B)$, 定义为:

$$P(A|B) = P(AB)/P(B).$$

独立 两个事件 A, B 若满足 $P(AB) = P(A)P(B)$, 则称 A, B 独立。

注1: 上式事实上也是独立的性质。(亦即 $P(A) = P(A|B)$.)

注2 (推广) : $P(A_{i_1}A_{i_2}\cdots) = P(A_{i_1})P(A_{i_2})\cdots, P(A_{i_1}) = P(A_{i_1}|A_{i_2}\cdots)$.
(任取有限个)

概率乘法定理 若干个**独立**事件之积的概率, 等于各事件的概率之乘积, 即

$$P(A_1A_2\cdots) = P(A_1)P(A_2)\cdots.$$

推论1 独立事件的任一部分也独立。

推论2 若一系列事件相互独立, 则将其中任一部分改为对立事件时, 所得事件列仍为相互独立。

完备事件群 设 B_1, B_2, \cdots 为有限或无限个事件, 它们两两互斥且在每次试验中至少发现一个, 用式表之, 即:

$$B_iB_j = \emptyset \quad (i \neq j),$$

$$B_1 + B_2 + \cdots = \Omega.$$

全概率公式

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots.$$

贝叶斯公式 在全概率公式的假定下, 有:

$$\begin{aligned} P(B_i|A) &= P(AB_i)/P(A) \\ &= P(B_i)P(A|B_i) / \sum_j P(B_j)P(A|B_j). \end{aligned}$$

2 随机变量及概率分布

2.1 一维随机变量

随机变量 “其值随机而定”的变量 *†

一种分类方式 离散型/连续型

2.1.1 离散型

概率函数 (离散型) 设 X 为离散型随机变量, 其全部可能值为 $\{a_1, a_2, \dots\}$, 则其概率函数为:

$$p_i = P(X = a_i).$$

注1: 显然有 $p_i \geq 0$, $p_1 + p_2 + \dots = 1$.

注2: 上式称为随机变量 X 的概率分布。

分布函数 (离散型) $F(x) = P(X \leq x) = \sum_{\{i|a_i \leq x\}} p_i$.

分布函数的性质

(1) $F(x)$ 单调非降。

(2) $x \rightarrow \infty, F(x) \rightarrow 1$; $x \rightarrow -\infty, F(x) \rightarrow 0$.

二项分布 若随机变量 X 的可能取值为 $0, 1, \dots, n$, 且概率分布为

$$P(X = i) = b(i; n, p) = \binom{n}{i} p^i (1-p)^{n-i} \quad (i = 0, 1, \dots, n),$$

则称 X 服从二项分布, 记为 $X \sim B(n, p)$.

*The 1st version of this chapter was completed on 2015.12.24

†Last edited on 2016.1.5

实际意义： 总抽样个数 n , 将废品个数 X 作为变量。

泊松分布 若随机变量 X 的可能取值为 $0, 1, \dots$, 且概率分布为

$$P(X = i) = e^{-\lambda} \lambda^i / i! \quad (i = 0, 1, \dots),$$

则称 X 服从泊松分布, 记为 $X \sim P(\lambda)$.

注： 若 $X \sim B(n, p)$, 其中 n 很大, p 很小, $np = \lambda$ 不太大时, X 的分布接近于泊松分布。(可做转化以简化计算)

超几何分布

$$P(X = m) = \binom{M}{m} \binom{N-M}{n-m} / \binom{N}{n} \quad (0 \leq m \leq M, n-m \leq N-M).$$

实际意义： 总个数 N , 废品个数 M , 抽样个数 n , 其中的废品数 X 作为变量。

负二项分布

$$P(X = i) = b(r-1; i+r-1, p)p = \binom{i+r-1}{r-1} p^r (1-p)^i \quad (i = 0, 1, \dots).$$

实际意义： 废品个数 r , 将发现第 r 个废品时总抽样次数 X 作为变量。

注： $r = 1$, $P(X = i) = p(1-p)^i \quad (i = 0, 1, \dots)$ 称为几何分布。

2.1.2 连续型

概率密度函数 (连续型) $f(x) = F'(x)$.

概率密度函数 (连续型) 的性质

(1) $f(x) \geq 0$.

(2) $\int_{-\infty}^{\infty} f(x)dx = 1.$

(3) 对于任何常数 $a < b$, 有:

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

注: 由于连续型变量取一个点的概率为0, 故区间断点存在与否无影响。

正态分布 如果一个随机变量具有概率密度函数

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

则称 X 为正态随机变量, 记为 $X \sim N(\mu, \sigma^2)$.

标准正态分布 $f(x) = e^{-\frac{x^2}{2}} / \sqrt{2\pi}$, $X \sim N(0, 1)$.

注1: 若 $X \sim N(\mu, \sigma^2)$, 则 $Y = (X - \mu) / \sigma \sim N(0, 1)$.

注2: 标准正态分布的分布函数为 $\Phi(x)$, 有 $\Phi(x) = 1 - \Phi(-x)$.

指数分布 如果一个随机变量 X 具有概率密度函数

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

则称 X 服从指数分布, 其中 $\lambda > 0$ 为参数。

分布函数

$$F(x) = \int_{-\infty}^x f(t)dt = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

均匀分布 如果一个随机变量 X 具有概率密度函数

$$f(x) = \begin{cases} 1/(b-a) & a \leq x \leq b \\ 0 & \text{其他} \end{cases}$$

则称 X 服从区间 $[a, b]$ 上的均匀分布, 并常记为 $X \sim R(a, b)$.

分布函数

$$F(x) = \begin{cases} 0 & x \leq a \\ (x-a)/(b-a) & a < x < b \\ 1 & x \geq b \end{cases}$$

2.2 多维随机变量 (随机向量)

n 维随机变量 一般地, 设 $X = (X_1, X_2, \dots, X_n)$ 为一个 n 维向量, 其每个分量都是一维随机变量。

2.2.1 离散型

概率函数 (离散型) 以 $\{a_{i1}, a_{i2}, \dots\}$ 记 X_i 的全部可能值 ($i = 1, 2, \dots$), 则事件 $\{X_1 = a_{1j_1}, \dots, X_n = a_{nj_n}\}$ 的概率

$$p(j_1, j_2, \dots, j_n) = P(X_1 = a_{1j_1}, \dots, X_n = a_{nj_n}) \\ (j_1 = 1, 2, \dots; \dots; j_n = 1, 2, \dots)$$

称为随机向量 $X = (X_1, X_2, \dots, X_n)$ 的概率函数或概率分布。

概率函数满足的条件

- (1) $p(j_1, j_2, \dots, j_n) \geq 0$.
- (2) $\sum_{j_n} \dots \sum_{j_2} \sum_{j_1} p(j_1, j_2, \dots, j_n) = 1$.

多项分布

$$P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = \frac{N!}{k_1! k_2! \dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}$$

(k_i 为非负整数, $k_1 + \dots + k_n = N$).

记为 $M(N, p_1, \dots, p_n)$.

2.2.2 连续型

概率密度函数 (连续型) 若 $f(x_1, \dots, x_n)$ 是定义在 \mathbb{R}^n 上的非负函数, 使对 \mathbb{R}^n 中的任意集合 A , 有

$$P(X \in A) = \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

则称 f 是 X 的概率密度函数。

二维正态分布

$$f(x_1, x_2) = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-a)^2}{\sigma_1^2} - \frac{2\rho(x_1-a)(x_2-b)}{\sigma_1\sigma_2} + \frac{(x_2-b)^2}{\sigma_2^2}\right)\right].$$

其中

$$-\infty < a, b < \infty, \sigma_1, \sigma_2 > 0, -1 < \rho < 1.$$

记为 $N(a, b, \sigma_1^2, \sigma_2^2, \rho)$.

有关注意事项

(1) 定义一维或者多位连续型随机变量时, 实质都在于有概率密度函数存在。(有密度函数的随机变量)

(2) 连续型随机变量不能简单定义为“其各分量都是一维连续型随机变量的那种随机向量”。

(3) 与一维一样, 也可以用概率分布函数去描述多维随机向量的概率分布, 其定义为:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

然而, 在多维情况下, 分布函数很少应用。

2.2.3 边缘分布

概念 设 $X = (X_1, \dots, X_n)$ 为一个 n 维随机向量。 X 有一定的分布 F , 这是一个 n 维分布。 X 的每个分量 X_i 都是一维随机变量, 都有各自的分布 F_i , 这些一维分布称为随机向量 X 或其分布 F 的“边缘分布”。

离散型 求和

多项分布 边缘分布为二项分布。

连续型 (n 维为例) 某分量 X_1 的概率密度函数

$$f_1(x_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \dots dx_n.$$

n 维正态分布 边缘分布为一维正态分布。

联合分布 将 X_1, \dots, X_n 作为一个有联系的整体来考虑, 把 (X_1, \dots, X_n) 的分布称为“联合分布”。

2.3 条件概率分布与随机变量的独立性

概念 在某种给定的条件之下 X 的概率分布。

2.3.1 离散型

导出 (以二维为例) 设 (X_1, X_2) 为一个二维离散型随机向量, X_1 的全部可能值为 a_1, a_2, \dots ; X_2 的全部可能值为 b_1, b_2, \dots ; 而 (X_1, X_2) 的联合概率分布为

$$p_{ij} = P(X_1 = a_i, X_2 = b_j) \quad (i, j = 1, 2, \dots).$$

依条件概率的定义, 有

$$P(X_1 = a_i | X_2 = b_j) = P(X_1 = a_i, X_2 = b_j) / P(X_2 = b_j) = p_{ij} / \sum_k p_{kj}.$$

2.3.2 连续型

定义 (以二维为例) 条件分布函数:

$$P(X_1 \leq x_1 | a \leq X_2 \leq b) = \int_{-\infty}^{x_1} dt_1 \int_a^b f(t_1, t_2) dt_2 / \int_a^b f_2(t_2) dt_2.$$

条件密度函数:

$$f(x_1 | a \leq X_2 \leq b) = \int_a^b f(x_1, t_2) dt_2 / \int_a^b f_2(t_2) dt_2.$$

推广

$$f(x_1, \dots, x_n) = g(x_1, \dots, x_k) h(x_{k+1}, \dots, x_n | x_1, \dots, x_k).$$

2.3.3 随机变量的独立性

定义 (连续型) 设 n 维随机变量 (X_1, \dots, X_n) 的联合密度函数为 $f(x_1, \dots, x_n)$, 而 X_i 的边缘密度函数为 $f_i(x_i)$ ($i = 1, \dots, n$). 如果

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n),$$

就称随机变量 X_1, \dots, X_n 相互独立, 简称独立。

结论1 如果连续变量 X_1, \dots, X_n 独立, 则对于任何 $a_i < b_i$ ($i = 1, \dots, n$), $A_i = a_i \leq X_i \leq b_i$ ($i = 1, \dots, n$) 这 n 个事件也独立。

结论2 如果连续变量 X_1, \dots, X_n 的概率密度函数满足 $f(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n)$, 则 X_1, \dots, X_n 独立, 且 $f_i(x_i)$ 与 $g_i(x_i)$ 只相差一个常数因子。

结论3 若 X_1, \dots, X_n 相互独立, 而 $Y_1 = g_1(X_1, \dots, X_m)$, $Y_2 = g_2(X_{m+1}, \dots, X_n)$, 则 Y_1 和 Y_2 独立。

定义 (离散型) 设 X_1, \dots, X_n 都是离散型随机变量。若对任何常数 a_1, \dots, a_n 都有

$$P(X_1 = a_1, \dots, X_n = a_n) = P(X_1 = a_1) \cdots P(X_n = a_n),$$

则称 X_1, \dots, X_n 相互独立。

示性函数

$$X_i = \begin{cases} 1, & \text{当事件} A \text{发生时} \\ 0, & \text{当事件} A \text{不发生时} \end{cases}$$

意指 X_i 的值“指示”了 A 是否发生。

2.4 随机变量的函数的概率分布

2.4.1 离散型

一般做法 把 $Y = g(X_1, \dots, X_n)$ 可以取的不同值找出来，把与某个值相应的全部 (X_1, \dots, X_n) 值的概率加起来即得。

常用结论

$$(1) \quad (X_1, \dots, X_n) \sim M(N; p_1, \dots, p_n), Y = X_1 + X_2 \sim B(N, p_1 + p_2).$$

$$(2) \quad X_1 \sim B(n_1, p), X_2 \sim B(n_2, p), Y = X_1 + X_2 \sim B(n_1 + n_2, p).$$

$$(3) \quad X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2), Y = X_1 + X_2 \sim P(\lambda_1 + \lambda_2).$$

2.4.2 连续型 (一般讨论)

设 (X_1, \dots, X_n) 有密度函数 $f(x_1, \dots, x_n)$, 而

$$Y_i = g_i(X_1, \dots, X_n) \quad (i = 1, \dots, n)$$

构成 (X_1, \dots, X_n) 到 (Y_1, \dots, Y_n) 的一一对应变换，其逆变换为

$$X_i = h_i(Y_1, \dots, Y_n) \quad (i = 1, \dots, n),$$

此变换的Jacobi行列式为

$$J(y_1, \dots, y_n) = \begin{vmatrix} \partial h_1 / \partial y_1 & \cdots & \partial h_1 / \partial y_n \\ \vdots & & \vdots \\ \partial h_n / \partial y_1 & \cdots & \partial h_n / \partial y_n \end{vmatrix},$$

则 (Y_1, \dots, Y_n) 的密度函数为

$$l(y_1, \dots, y_n) = f(h_1, \dots, h_n) |J(y_1, \dots, y_n)|.$$

常用结论

(1) $Y \sim X^2, l(y) = \frac{1}{2}y^{-1/2}(f(\sqrt{y}) + f(-\sqrt{y}))$.

(2) $X \sim N(\mu, \sigma^2), aX + b \sim N(a\mu + b, a^2\sigma^2)$.

2.4.3 随机变量和的密度函数

一般定义 (二维) 设 (X_1, X_2) 的联合密度函数为 $f(x_1, x_2)$, $Y = X_1 + X_2$ 的密度函数为

$$l(y) = \int_{-\infty}^{\infty} f(y-x, x)dx = \int_{-\infty}^{\infty} f(x, y-x)dx.$$

如果 X_1, X_2 独立, 上式可改为

$$l(y) = \int_{-\infty}^{\infty} f_1(y-x)f_2(x)dx = \int_{-\infty}^{\infty} f_1(x)f_2(y-x)dx.$$

常用结论

(1) X_1, X_2 不独立, 其联合分布为 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则 $Y = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$.

(2) 若 X_1, \dots, X_n 相互独立, 分别服从正态分布 $N(\mu_1, \sigma_1^2), \dots, N(\mu_n, \sigma_n^2)$, 则 $X_1 + \dots + X_n$ 服从正态分布 $N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$.

卡方分布 若 $n > 0$, 则

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

是概率密度函数。记为 χ_n^2 .

常用结论

(1) 若 X_1, \dots, X_n 独立, 且都服从正态分布 $N(0, 1)$, 则

$$Y = X_1^2 + \dots + X_n^2 \sim \chi_n^2.$$

(2) 设 X_1, X_2 独立, $X_1 \sim \chi_m^2, X_2 \sim \chi_n^2$, 则 $X_1 + X_2 \sim \chi_{m+n}^2$.

(3) 若 X_1, \dots, X_n 独立, 且都服从指数分布, 则

$$X = 2\lambda(X_1 + \dots + X_n) \sim \chi_{2n}^2.$$

2.4.4 随机变量商的密度函数

一般求解公式 设 (X_1, X_2) 有密度函数 $f(x_1, x_2)$, $Y = X_2/X_1$. (X_1 只取正值的情况), Y 的密度函数为

$$l(y) = \int_0^{\infty} x_1 f(x_1, x_1 y) dx_1.$$

若 X_1, X_2 独立, 则上式成为

$$l(y) = \int_0^{\infty} x_1 f_1(x_1) f_2(x_1 y) dx_1.$$

t 分布 设 X_1, X_2 独立, $X_1 \sim \chi_n^2, X_2 \sim N(0, 1)$, $Y = X_2/\sqrt{X_1/n}$, 则

$$Y \sim t_n,$$

其中 Y 的密度函数为

$$t_n(y) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}.$$

F 分布 设 X_1, X_2 独立, $X_1 \sim \chi_n^2$, $X_2 \sim \chi_m^2$, $Y = m^{-1}X_1/(n^{-1}X_2)$, 则

$$Y \sim F_{mn},$$

其中 Y 的密度函数为

$$f_{mn}(y) = m^{m/2}n^{n/2} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} y^{m/2-1} (my+n)^{-(m+n)/2} \quad (y > 0).$$

当 $y \leq 0$ 时 $f_{mn}(y) = 0$, 因为 Y 只取正值。

三大分布的几条重要性质 变量 X_1, \dots, X_n iid., $\sim N(\mu, \sigma^2)$. 记 $\bar{X} = \sum_{i=1}^n X_i/n$, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$.

$$(1) \quad \sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1).$$

$$(2) \quad (n-1)S^2/\sigma^2 = \sum_{i=1}^n (X_i - \bar{X})^2/\sigma^2 \sim \chi_{n-1}^2.$$

$$(3) \quad \sqrt{n}(\bar{X} - \mu)/S \sim t_{n-1}.$$

(4) 中 X_1, \dots, X_n iid., $\sim N(\mu_1, \sigma_1^2)$, Y_1, \dots, Y_m iid., $\sim N(\mu_2, \sigma_2^2)$, 其余与前三条记号相同。

(4)

$$\left[\sum_{j=1}^m (Y_j - \bar{Y})^2 / (\sigma_2^2(m-1)) \right] / \left[\sum_{i=1}^n (X_i - \bar{X})^2 / (\sigma_1^2(n-1)) \right] \sim F_{m-1, n-1}.$$

若 $\sigma_1^2 = \sigma_2^2$, 则

$$\frac{\sqrt{\frac{nm(n+m-2)}{n+m}} [(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)]}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2 \right]^{1/2}} \sim t_{n+m-2}.$$

3 随机变量的数字特征

3.1 数学期望 (均值) 与中位数

3.1.1 数学期望

定义 (离散, 有限) 设随机变量 X 只取有限个可能值 a_1, \dots, a_m , 其概率分布为 $P(x = a_i) = p_i$ ($i = 1, \dots, m$). 则 X 的数学期望 (也常称均值), 记为 $E(X)$ 或 EX , 定义为^{*†}

$$E(X) = a_1 p_1 + \dots + a_m p_m.$$

定义 (离散, 无限) 如果

$$\sum_{i=1}^{\infty} |a_i| p_i < \infty,$$

则称

$$E(X) = \sum_{i=1}^{\infty} a_i p_i < \infty.$$

定义 (连续) 设 X 有概率密度函数 $f(x)$, 如果

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

则称

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

性质

$$(1) \quad E(X_1 + \dots + X_N) = E(X_1) + \dots + E(X_n).$$

$$(2) \quad X_i \text{ 独立, } E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n).$$

^{*}The 1st version of this chapter was completed on 2015.12.25

[†]Last edited on 2016.1.6

$$(3) \quad E(g(X)) = \sum_i g(a_i) p_i \quad (\text{当 } \sum_i |g(a_i)| p_i < \infty \text{ 时})$$

或

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (\text{当 } \int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty \text{ 时})$$

常用结论

$$(1) \quad X \sim B(n, p), \quad E(X) = np.$$

$$(2) \quad X \sim P(\lambda), \quad E(X) = \lambda.$$

$$(3) \quad X \sim \chi_n^2, \quad E(X) = n.$$

$$(4) \quad X \sim t_n, \quad E(X) = 0 \quad (n > 1 \text{ 否则不满足定义条件}).$$

$$(5) \quad X \sim F_{m,n}, \quad E(X) = n/(n-2) \quad (n > 2 \text{ 否则不满足定义条件}).$$

条件数学期望

$$E(Y) = E[E(Y|x)|_{x=X}] = E[E(Y|X)].$$

$$E(Y|x) = \int_{-\infty}^{\infty} y f(y|x) dy \quad (\text{以二维连续情况为例}).$$

中位数 设连续型随机变量 X 的分布函数为 $F(x)$, 则满足条件

$$P(x \leq (<, \geq, >) m) = F(m) = 1/2$$

的数 m 被称为 X 或分布 F 的中位数。

注1: 总存在, 但可以不唯一。

注2: 离散型的类似定义不是很理想。

3.2 方差与矩

方差 设 X 为随机变量, 分布为 F , 则

$$\text{Var}(X) = E(X - EX)^2 = E(X^2) - EX^2$$

称为 X 或分布 F 的方差。

标准差 其平方根 $\sqrt{\text{Var}(X)}$ (取正值) 称为 X 或分布 F 的标准差。

方差的性质

- (1) 若 c 为常数 (下同), 则 $\text{Var}(c) = 0$.
- (2) $\text{Var}(X + c) = \text{Var}(X)$.
- (3) $\text{Var}(cX) = c^2 \text{Var}(X)$.
- (4) X_i 独立, 则 $\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$.

常用结论

- (1) $X \sim B(n, p)$, $\text{Var}(X) = np(1 - p)$.
- (2) $X \sim P(\lambda)$, $\text{Var}(X) = \lambda$.
- (3) 指数分布的方差为 $1/\lambda^2$.
- (4) 均匀分布的方差为 $(b - a)^2/12$.
- (5) $X \sim N(\mu, \sigma^2)$, $\text{Var}(X) = \sigma^2$.
- (6) $X \sim \chi_n^2$, $\text{Var}(X) = 2n$.

(7) $X \sim t_n, \text{Var}(X) = n/(n-2)$ ($n > 2$ 否则不满足定义条件).

(8) $X \sim F_{m,n}, \text{Var}(X) = 2n^2(m+n-2)/[m(n-2)^2(n-4)]$ ($n > 4$ 否则不满足定义条件).

(9) (标准化) $E(X) = a, \text{Var}(X) = \sigma^2; Y = (X-a)/\sigma, E(Y) = 0, \text{Var}(Y) = 1.$

矩 设 X 为随机变量, c 为常数, k 为正整数. 则量 $E[(X-c)^k]$ 称为 X 关于 c 点的 k 阶矩.

(1) $c = 0$, 这时 $\alpha_k = E(X^k)$ 称为 X 的 k 阶原点矩.

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

矩的应用

(1) “偏度系数” $\beta_1 = \mu_3/\mu_2^{3/2}.$

(2) “峰度系数” $\beta_2 = \mu_4/\mu_2^2.$

3.3 协方差与相关系数

记 $E(X) = m_1, E(Y) = m_2, \text{Var}(X) = \sigma_1^2, \text{Var}(Y) = \sigma_2^2.$

协方差 称 $E[(X-m_1)(Y-m_2)]$ 为 X, Y 的协方差, 并记为 $\text{Cov}(X, Y).$

协方差的性质

(1) 若 X, Y 独立, 则 $\text{Cov}(X, Y) = 0.$ 反过来不一定成立.

(2) $[\text{Cov}(X, Y)] \leq \sigma_1^2 \sigma_2^2.$ 等号当且仅当 X, Y 之间有严格线性关系.

$$(3) \text{Cov}(c_1X + c_2, c_3X + c_4) = c_1c_3 \text{Cov}(X, Y).$$

$$(4) \text{Cov}(X, Y) = E(X, Y) - m_1m_2.$$

相关系数 称 $\text{Cov}(X, Y)/(\sigma_1\sigma_2)$ 为 X, Y 的相关系数, 并记为 $\text{Corr}(X, Y)$.

相关系数的性质

(1) 若 X, Y 独立, 则 $\text{Corr}(X, Y) = 0$.

(2) $-1 \leq \text{Cov}(X, Y) \leq 1$. 等号当且仅当 X, Y 之间有严格线性关系。

相关系数性质的几点解释

(1) 不相关不等于独立。

(2) 相关系数只刻画了 X, Y 之间“线性”关系的程度。(对于正态分布的情况则是完美的刻画)

(3) 若 $0 < |\text{Cov}(X, Y)| < 1$, 则可以说 X, Y 之间有“一定程度的”线性关系。

(4) “线性相关”的意义还可以从最小二乘法的角度去解释。

3.4 大数定理和中心极限定理

大数定理 设 X_1, \dots, X_n, \dots iid., 记它们的公共均值为 a , 又设它们的方差存在并记为 σ^2 . 则对任意给定的 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - a| \geq \varepsilon) = 0.$$

(“ \bar{X}_n 依概率收敛于 a ”)

一个重要特例

$$\lim_{n \rightarrow \infty} P(|p_n - p| \geq \varepsilon) = 0.$$

马尔科夫不等式 若 Y 为只取非负值的随机变量, 则对任意给定的 $\varepsilon > 0$, 有

$$P(Y \geq \varepsilon) \leq E(Y)/\varepsilon.$$

切比雪夫不等式 若 $\text{Var}(Y)$ 存在, 则

$$P(|Y - EY| \geq \varepsilon) \leq \text{Var}(Y)/\varepsilon^2.$$

中心极限定理 设 X_1, \dots, X_n, \dots iid., 记它们的公共均值为 a , 又设它们的方差有限且非零并记为 σ^2 . 则对任意实数 x , 有

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{\sqrt{n}\sigma}(X_1 + \dots + X_n - na) \leq x\right) = \Phi(x).$$

两点分布情况

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{\sqrt{np(1-p)}}(X_1 + \dots + X_n - np) \leq x\right) = \Phi(x).$$

两点分布情况的修正

如果 $t_1 < t_2$ 是两个正整数, 则在 n 相当大的情况下有

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

其中

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

进行如下修正

$$y_1 = \left(t_1 - \frac{1}{2} - np\right) / \sqrt{np(1-p)},$$

$$y_2 = \left(t_2 + \frac{1}{2} - np\right) / \sqrt{np(1-p)},$$

一般可以提高精度。

4 参数估计

4.1 数理统计学的基本概念

数理统计学 使用概率论和数学的方法，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），以对所研究的问题做出推断（称为统计推断）。*†

总体 与所研究的问题有关的对象（个体）的全体所构成的集合。

样本 按一定的规定从总体中抽出的一部分个体。

统计量 完全由样本决定的量。

4.2 矩估计、极大似然估计

点估计 用一个点去估计另一个点。

4.2.1 矩估计法

样本原点矩

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

样本中心矩

$$m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k.$$

一般步骤

(1) 求出总体矩 α_k, μ_k 与（带未知参数 θ_k 的）总体分布的关系式；

*The 1st version of this chapter was completed on 2015.12.27

†Last edited on 2016.1.7

(2) 反解出总体分布的未知参数 θ_k ;

(3) 用样本矩代替总体矩得到未知参数的估计值 $\hat{\theta}_k$.

原则 能用低阶矩处理的就不用高阶矩。

4.2.2 极大似然估计法

一般步骤

(1) 设总体有分布 $f(x; \theta_1, \dots, \theta_k)$, X_1, \dots, X_n 为样本, 则样本 (X_1, \dots, X_n) 的分布 (概率密度函数) 为

$$L(x_1, \dots, x_n; \theta_1, \dots, \theta_k) = f(x_1; \theta_1, \dots, \theta_k) \cdots f(x_n; \theta_1, \dots, \theta_k).$$

(2) 采用满足条件

$$L(X_1, \dots, X_n; \theta_1^*, \dots, \theta_k^*) = \max_{\theta_1, \dots, \theta_k} L(X_1, \dots, X_n; \theta_1, \dots, \theta_k).$$

的 $(\theta_1^*, \dots, \theta_k^*)$ 作为未知参数的估计值。

(3) 计算 L 最大值的一种方法 (似然方程组)

$$\frac{\partial \ln L}{\partial \theta_i} = 0 (i = 1, \dots, k).$$

(如果该方程组无法应用则需要从定义出发。)

4.3 点估计的优良性准则

$$E_{\theta_1, \dots, \theta_k} [\hat{g}(X_1, \dots, X_n)] = g(\theta_1, \dots, \theta_k),$$

则称 \hat{g} 是 $g(\theta_1, \dots, \theta_k)$ 的一个无偏估计量。

含义 没有系统误差, 但随机误差总是存在。

注：矩估计和极大似然估计不一定是无偏估计量！

例 m_2 不是总体方差 σ^2 的无偏估计，而样本方差 $S^2 = \frac{n}{n-1}m_2$ 是。

4.3.1 最小方差无偏估计

均方误差

$$M_{\hat{\theta}}(\theta) = E_{\theta}[\hat{\theta}(X_1, \dots, X_n) - \theta]^2 = \text{Var}_{\theta}(\hat{\theta}) + [E_{\theta}(\hat{\theta}) - \theta]^2.$$

最小方差无偏估计 (MVU估计) 设 $\hat{\theta}$ 为 $g(\theta)$ 的无偏估计。若对其任一
无偏估计 $\hat{\theta}_1$, 都有

$$\text{Var}_{\theta}(\hat{\theta}) \leq \text{Var}_{\theta}(\hat{\theta}_1)$$

对 θ 的任意可能取的值都成立，则称 $\hat{\theta}$ 为 $g(\theta)$ 的一个MVU估计。

4.4 区间估计

区间估计 用一个区间去估计未知参数，即把未知参数值估计在某两个界限之间。

置信系数 给定一个很小的数 $\alpha > 0$. 如果对于参数 θ 的任何值，都有

$$P_{\theta}(\hat{\theta}_1(X_1, \dots, X_n) \leq \theta \leq \hat{\theta}_2(X_1, \dots, X_n)) = 1 - \alpha,$$

则称区间估计 $[\hat{\theta}_1, \hat{\theta}_2]$ 的置信系数为 $1 - \alpha$.

4.4.1 枢轴变量法

一般步骤

(1) 找一个与要估计的参数 $g(\theta)$ 有关的统计量 T , 一般是其一个良好的点估计;

(2) 设法找出某一函数 $S(T, g(\theta))$, 其分布 F 要与 θ 无关, S 称为枢轴变量;

(3) 对于任意 $a < b$, $a \leq S(T, g(\theta)) \leq b \Rightarrow A \leq g(\theta) \leq B$. 其中 A, B 与 θ 无关;

(4) 取分布 F 的上 $\alpha/2$ 分位点 $w_{\alpha/2}$ 和上 $1 - \alpha/2$ 分位点 $w_{1-\alpha/2}$, 令其分别为 a, b , 解出 A, B , $[A, B]$ 即为待求区间。

常用实例

(1) 从正态总体 $N(\mu, \sigma^2)$ 中抽取样本 X_1, \dots, X_n , μ, σ^2 都未知, μ 的区间估计(一样本 t 区间估计)为

$$[\bar{X} - St_{n-1}(\alpha/2)/\sqrt{n}, \bar{X} + St_{n-1}(\alpha/2)/\sqrt{n}];$$

σ^2 的区间估计为

$$[(n-1)S^2/\chi_{n-1}^2(\alpha/2), (n-1)S^2/\chi_{n-1}^2(1-\alpha/2)].$$

(2) 两个正态总体, 分布为 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$. 从中分别抽样 $X_1, \dots, X_n, Y_1, \dots, Y_m$, $\mu_1, \mu_2, \sigma_1^2 = \sigma_2^2$ 都未知, $\mu_1 - \mu_2$ 的区间估计(两样本 t 区间估计)为

$$[(\bar{X} - \bar{Y}) - St_{n+m-2}(\alpha/2)\sqrt{\frac{n+m}{nm}}, (\bar{X} - \bar{Y}) + St_{n+m-2}(\alpha/2)\sqrt{\frac{n+m}{nm}}];$$

若 σ_1^2 与 σ_2^2 不一定相等, σ_1^2/σ_2^2 的区间估计为

$$[(S_1^2/S_2^2)F_{m-1, n-1}(1-\alpha/2), (S_1^2/S_2^2)F_{m-1, n-1}(\alpha/2)].$$

(3) 设 X_1, \dots, X_n 为取自指数分布的样本, 其参数 λ 的估计为

$$[\chi_{2n}^2(1-\alpha/2)/(2n\bar{X}), \chi_{2n}^2(\alpha/2)/(2n\bar{X})]$$

其总体均值 $1/\lambda$ 的估计为

$$[2n\bar{X}/\chi_{2n}^2(\alpha/2), 2n\bar{X}/\chi_{2n}^2(1-\alpha/2)].$$

注 更详细一(两)样本构造枢轴变量的方法的请参照P31-32的两个表格。

常用结论 设 θ 的 $1 - \alpha$ 区间估计为 $[\hat{\theta}_1, \hat{\theta}_2]$, 且 $\theta = f(t)$, 其中 $f(t)$ 为给定的连续单调函数, 则 t 的 $1 - \alpha$ 区间估计为 $[f_{-1}(\hat{\theta}_1), f_{-1}(\hat{\theta}_2)]$.

4.4.2 大样本法

概念 利用**极限分布**, 主要是中心极限定理 (正态分布), 以建立枢轴变量。

常用实例

(1) $Y_n \sim B(n, p)$, 求 p 的区间估计, 近似有

$$(Y_n - np) / \sqrt{np(1-p)} \sim N(0, 1). \text{(区间端点满足 } = u_{\alpha/2}^2 \text{, 下同)}$$

(2) 设 X_1, \dots, X_n 为取自泊松分布 $P(\lambda)$ 的样本, $Y_n = X_1 + \dots + X_n$, 求 λ 的区间估计, 近似有

$$(Y_n - n\lambda)^2 / (n\lambda) \sim N(0, 1).$$

(3) 设某总体有均值 θ , 方差 σ^2 且均未知, 抽取样本 X_1, \dots, X_n , 求 θ 的区间估计, 近似有

$$\sqrt{n}(\bar{X} - \theta) / S \sim N(0, 1).$$

(4) 两个正态总体, 分布为 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$. 从中分别抽样 $X_1, \dots, X_n, Y_1, \dots, Y_m, \mu_1, \mu_2, \sigma^2$ 都未知, 求 $\mu_1 - \mu_2$ 的区间估计, 近似为

$$[(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)] / \sqrt{S_1^2/n - S_2^2/m} \sim N(0, 1).$$

4.4.3 置信界

定义 设 X_1, \dots, X_n 是从某一总体抽出来的样本, 总体分布包含未知参数 $\theta, \bar{\theta} = \bar{\theta}(X_1, \dots, X_n), \underline{\theta} = \underline{\theta}(X_1, \dots, X_n)$ 都是统计量 (与 θ 无关), 则

(1) 若对 θ 一切可取的值, 有

$$P_{\theta}(\bar{\theta}(X_1, \dots, X_n) \geq \theta) = 1 - \alpha,$$

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

(2) 若对 θ 一切可取的值, 有

$$P_{\theta}(\underline{\theta}(X_1, \dots, X_n) \leq \theta) = 1 - \alpha,$$

则称 $\underline{\theta}$ 为 θ 的一个置信系数为 $1 - \alpha$ 的置信下界。

5 假设检验

5.1 问题提法和基本概念

原假设（零假设）和对立假设 在假设检验中，常把一个被检验的假设叫做原假设，而其对立面就叫做对立假设（可以指全体也可以指一个或一些特殊情况）。*†

检验统计量 在检验一个假设时所使用的统计量称为检验统计量。

接受域、拒绝域（否定域、临界域）和临界值 使原假设得到接受的那些样本所在的区域被称为该检验的接受域；而使原假设被否定的那些样本所在的区域成为该检验的拒绝域。两者的界限成为临界值。

简单假设和复合假设 不论是原假设还是对立假设，若其中只含有一个参数值，则成为简单假设；否则就称为复合假设。

功效函数 设总体分布包含若干个未知参数 $\theta_1, \dots, \theta_k$, H_0 是关于这些参数的一个原假设，设有了样本 X_1, \dots, X_n , 而 Φ 是基于这些样本而对 H_0 所做的一个检验。则称检验 Φ 的功效函数为

$$\beta_{\Phi}(\theta_1, \dots, \theta_k) = P_{\theta_1, \dots, \theta_k}(\text{在检验 } \Phi \text{ 之下, } H_0 \text{ 被否定}),$$

它是未知参数 $\theta_1, \dots, \theta_k$ 的函数。

两类错误

第I类错误 H_0 正确，但被否定了。（弃真）

第II类错误 H_0 不正确，但被接受了。（存伪）

*The 1st version of this chapter was completed on 2015.12.29

†Last edited on 2016.1.7

显著性水平 设 Φ 是原假设 H_0 的一个检验, $\beta_{\Phi}(\theta_1, \dots, \theta_k)$ 为其功效函数, α 为常数 ($0 \leq \alpha \leq 1$). 如果

$$\beta_{\Phi}(\theta_1, \dots, \theta_k) \leq \alpha \quad (\text{对任何 } (\theta_1, \dots, \theta_k) \in H_0),$$

则称 Φ 是 H_0 的一个显著性水平为 α 的检验。

假设检验的原则

- (1) 将受保护的對象置为零假设。
- (2) 如果你希望“证明”某个命题, 就取相反结论或者其中一部分作为零假设。

假设检验的步骤

- (1) 求出未知参数 θ 的一个较优的点估计 $\hat{\theta}$.
- (2) 以 $\hat{\theta}$ 为基础, 寻找一个检验统计量 $T = t(X_1, \dots, X_n)$ 且使得 $\theta = \theta_0$ 时, T 的分布已知, 从而容易通过查表或计算得到这个分布的分位数, 用以作为检验的临界值。
- (3) 以检验统计量 T 为基础, 根据对立假设 H_1 的实际意义, 寻找适当形状的拒绝域, 其中包含一个到两个临界值。
- (4) 当零假设成立时, 犯第I类错误的概率小于等于给定的显著性水平 α , 这给出了一个关于临界值的方程, 解出临界值, 其等于 T 的分位数, 这即确定了检验的拒绝域。
- (5) 如果给出样本观测值, 则可算出检验统计量的样本观测值, 如落在拒绝域则可拒绝零假设, 否则不能。

5.2 重要参数检验

5.2.1 一样本正态总体 $N(\mu, \sigma^2)$ 的假设检验

表 6.2.1 一样本正态总体 $N(\mu, \sigma^2)$ 的假设检验

检验对象	检验统计量	分布	拒绝域
$\mu(\sigma^2\text{已知})$	$U = \sqrt{n}(\bar{X} - \mu_0)/\sigma$	$N(0,1)$	$ U > u_{\alpha/2}$
			$U > u_{\alpha}$
			$U < -u_{\alpha}$
$\mu(\sigma^2\text{未知})$	$T = \sqrt{n}(\bar{X} - \mu_0)/S$	t_{n-1}	$ T > t_{n-1}(\frac{\alpha}{2})$
			$T > t_{n-1}(\alpha)$
			$T < -t_{n-1}(\alpha)$
$\sigma^2(\mu\text{已知})$	$\chi^2 = \sum_{i=1}^n (X_i - \mu)^2 / \sigma_0^2$	χ_n^2	$\chi^2 > (<) \chi_n^2(1 - \frac{\alpha}{2})$
			$\chi^2 > \chi_n^2(\alpha)$
			$\chi^2 > \chi_n^2(1 - \alpha)$
$\sigma^2(\mu\text{未知})$	$\chi^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / \sigma_0^2$	χ_{n-1}^2	$\chi^2 > (<) \chi_{n-1}^2(1 - \frac{\alpha}{2})$
			$\chi^2 > \chi_{n-1}^2(\alpha)$
			$\chi^2 > \chi_{n-1}^2(1 - \alpha)$

注:

有关均值的检验: 对立假设分别为 $\mu \neq \mu_0$, $\mu > \mu_0$ 和 $\mu < \mu_0$.

有关方差的检验: 对立假设分别为 $\sigma^2 \neq \sigma_0^2$, $\sigma^2 > \sigma_0^2$ 和 $\sigma^2 < \sigma_0^2$.

*** 双侧都是 $\alpha/2$, 单侧都是 α

5.2.2 两样本正态总体的假设检验

表 6.2.2 两样本正态总体的假设检验

检验对象	检验统计量	分布	拒绝域
均值(方差已知)	$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$	$N(0, 1)$	$ U > u(\alpha/2)$
			$U > u(\alpha)$
			$U < -u(\alpha)$
均值(方差未知*)	$T = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{1/m + 1/n}}$	t_{m+n-2}	$ T > t_{m+n-2}(\alpha/2)$
			$T > t_{m+n-2}(\alpha)$
			$T < -t_{m+n-2}(\alpha)$
方差(均值已知)	$F = \frac{\sum_{i=1}^m (X_i - \mu_1)^2 / m}{\sum_{i=1}^n (X_i - \mu_2)^2 / n}$	$F_{m,n}$	$F > F_{m,n}(\alpha/2) \text{ or } F < \frac{1}{F_{n,m}(\alpha/2)}$
			$F > F_{m,n}(\alpha)$
			$F < \frac{1}{F_{n,m}(\alpha)}$
方差(均值未知)	$F = \frac{S_1^2}{S_2^2}$	$F_{m-1,n-1}$	$F > F_{m-1,n-1}(\alpha/2) \text{ or } F < \frac{1}{F_{n-1,m-1}(\alpha/2)}$
			$F > F_{m-1,n-1}(\alpha)$
			$F < \frac{1}{F_{n-1,m-1}(\alpha)}$

注:

有关均值的检验: 对立假设分别为 $\mu \neq \mu_0$, $\mu > \mu_0$ 和 $\mu < \mu_0$.

有关方差的检验: 对立假设分别为 $\sigma_1^2 \neq \sigma_2^2$, $\sigma_1^2 > \sigma_2^2$ 和 $\sigma_1^2 < \sigma_2^2$.

* 假定方差相等。

5.2.3 成对数据

(1) 是同一个个体的两个指标, 故具有很大的相关性而绝对不是独立的。

(2) 两样本检验要求 X_1, \dots, X_m 是同分布的 (Y_1, \dots, Y_n 亦然), 而成对数据无此要求, 只要求 $X_1 - Y_1, \dots, X_n - Y_n$ 是同分布的。

(3) 实际问题中如果发现有两个样本且样本量相等, 则要检查独立性和同分布性, 否则可能是成对数据。

5.2.4 二项分布中未知参数 p 的假设检验

设 (X_1, \dots, X_n) 是取自总体 $X \sim B(n, p)$ 的一个样本。常见的假设有三种:

$$(1) H_0 : p = p_0 \leftrightarrow H_1 : p \neq p_0;$$

$$(2) H_0 : p = (\leq) p_0 \leftrightarrow H_1 : p > p_0;$$

$$(3) H_0 : p = (\geq) p_0 \leftrightarrow H_1 : p < p_0.$$

假设样本量 n 较大, 取显著性水平为 α , p 的极大似然估计为 \bar{X} , 取“标准化”的检验统计量

$$T = \frac{n(\bar{X} - p_0)}{\sqrt{np_0(1 - p_0)}}$$

由中心极限定理知 H_0 成立时近似有 $T \sim N(0, 1)$. 于是上述三种检验的拒绝域分别为

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

5.3 拟合优度检验

5.3.1 离散分布且理论分布完全已知

设有一总体 X . 根据经验得知 X 的分布为

$$H_0 : P(X = a_i) = p_i \quad (i = 1, \dots, k),$$

其中 a_1, p_i 已知, 且 a_i 两两不同, $p_i > 0$. 现从中抽样 n 次, 落在 a_i 的观测数为 n_i . 根据大数定理, 在零假设成立时, 理论频数 np_i 与观测频数 n_i 接近. 而检验统计量取为

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \sum \frac{(\text{观测频数} - \text{理论频数})^2}{\text{理论频数}}$$

可以证明, 如果零假设成立, 则在样本大小趋近于无穷大时, 检验统计量的极限分布为 χ_{k-1}^2 .

5.3.2 离散分布且理论分布含若干未知参数

应该用适当的估计如极大似然估计来代替这些参数以得到 p_i 的估计 \hat{p}_i , 得到的统计量记为

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

在零假设下, 其极限分布为 χ_{k-1-r}^2 , 其中 r 为估计的独立参数。

5.3.3 离散分布列联表的独立性检验

列联表 一种按照两个属性作双向分类的表。一般地, 如果第一个属性有 a 个水平, 第二个属性有 b 个水平, 称为 $a \times b$ 表。

独立性检验问题中, 零假设为

$$H_0: \text{属性 } A \text{ 与属性 } B \text{ 独立.}$$

假设样本量为 n , 第 (i, j) 格频数为 n_{ij} . 记 $p_{ij} = P(\text{属性 } A, B \text{ 分别处于水平 } i, j)$, $u_i = P(\text{属性 } A \text{ 有水平 } i)$, $v_j = P(\text{属性 } B \text{ 有水平 } j)$. 其极大似然估计为其频率, 即

$$\hat{u}_i = \frac{n_{i.}}{n} = \sum_{j=1}^b \frac{n_{ij}}{n}, \quad \hat{v}_j = \frac{n_{.j}}{n} = \sum_{i=1}^a \frac{n_{ij}}{n}.$$

故取检验统计量为

$$\chi^2 = \sum_{i=1}^a \sum_{j=1}^b \frac{(n_{ij} - n_i \cdot n_{.j} / n)^2}{(n_i \cdot n_{.j} / n)}.$$

在零假设下, 其极限分布为 $\chi_{(a-1)(b-1)}^2$.

5.3.4 离散分布列联表的齐一性检验

检验某一个属性 A 的各个水平对应的另一属性的 B 的分布全部相同, 虽然与独立性检验有着本质上的区别 (A 是非随机的), 但是所采用的检验方法和独立性检验完全一样。

5.3.5 连续总体情形

连续情况的假设可以通过适当的离散化总体分布, 采用拟合优度法来做检验 (分成 k 个子区间)。其余方法均与离散情况的相同。

注： 使用 χ^2 进行拟合优度检验一般要求 $n \geq 50, n\hat{p}_j \geq 5$. 如果不满足这个条件，最好把某些组作适当合并。

6 Acknowledgement

感谢丛聪同学协助校对本提纲，在此预祝他概统考出4.3的好成绩！

7 References

- [1] 《概率论与数理统计》陈希孺，中国科学技术大学出版社，2009.2
- [2] 《概率论与数理统计讲义》中国科学技术大学统计与金融系概率统计教研室，2008.4