

第一篇 概率论

第一章 随机事件和概率

第一节 事件

在个别试验中其结果没有规律性, 在大量重复试验中其结果又具有统计规律性的现象称为**随机现象**。对随机现象加以研究所进行的观察或试验, 称为**随机试验**。

随机试验 E 的所有可能的结果组成的**集合**称为 E 的**样本空间**, 记为 Ω 。 E 中的每一结果, 称为集合中的一个元素, 也称为一个**样本点**。样本空间 Ω 的任意子集称为 E 的**随机事件**, 简称**事件**。每一样本点称为**基本事件**。如掷一颗骰子, 样本空间 $\Omega = \{1, 2, 3, 4, 5, 6\}$; $1, 2, 3, 4, 5, 6$ 都是样本点; $S = \{1, 2, 3\}$ 是事件, $A = \{1\}$ 等为基本事件, 所有可能的事件为 $C_6^1 + C_6^2 + C_6^3 + C_6^4 + C_6^5 + C_6^6 = 63$ 种, 称为 E 中所有可能的**随机事件定义域** F 。在每次试验中, 当且仅当某一子集中的一个样本点出现, 称对应子集的事件发生; Ω 是自身的子集, 称为**必然事件**; $\Phi = \{ \}$ 不包含任何样本点, 称为**空集**, 又称为**不可能事件**。 $B = \{7\}$ 也是不可能事件。

第二节 利用文氏图直观理解和推导事件的关系

概率论中的文氏图分为**集合文氏图**和**概率文氏图**两种。

一、集合文氏图

在一个矩形区域 (表示样本空间 Ω) 内, 根据所给事件的集合关系画一系列圆, 圆围成的平面区域表示一个集合, 即事件, 并且在需要时把集合元素写在区域内, 从而将集合的交、并、余等集合关系用几何图形予以表示, 并在此基础上利用面积的加减或位置关系分析事件关系或概率问题。**集合文氏图**既可分析事件的集合关系, 又可以分析事件的概率关系。比如: $AB = \Phi \Rightarrow P(AB) = 0$ 。

二、事件的 8 个基本运算关系

事件既然是集合, 因此, 完全遵循集合的逻辑运算法则。

①包含: $A \subset B$, 即 A 事件发生必导致 B 事件发生。且有 $A \subset B \Leftrightarrow \bar{A} \supset \bar{B}$ 。

②相等: $A = B$, 即 $A \subset B$ 且 $B \subset A$ 。

③和 (并): $A \cup B$ 或 $A + B$ (注意此处的 $+$ 号只是 $\cup$ 的简化表示, 不一定满足数字运算的 $+$ 号的功能), 即 A 和 B 至少有一事件发生。如果 $A \cup B = B$, 则等价结论有 $\Leftrightarrow A \subset B \Leftrightarrow \bar{A} \supset \bar{B} \Leftrightarrow \bar{A}\bar{B} = \Phi$

④互逆或对立: $A \cup B = \Omega, A \cap B = \Phi \Leftrightarrow \bar{A} \cup \bar{B} = \Omega, \bar{A} \cap \bar{B} = \Phi \Leftrightarrow AB = \bar{A}\bar{B}$, 称 A, B (或 $\bar{A}, \bar{B}$) 为互逆或对立事件。

⑤互斥或不相容: $A \cap B = \Phi \Leftrightarrow \bar{A}\bar{B} = A$, 称为互斥或不相容事件, 注意与对立事件的区别和联系。

A, B 不相容不能推出 $\bar{A}\bar{B}$, 因为 $AB = \Phi \not\Rightarrow \bar{A}\bar{B} = \overline{A+B} = \Phi$ 。

⑥积 (交): $A \cap B = AB$, 即 A 和 B 同时发生。

⑦差: $A - B = A\bar{B}$, 称为差事件, 即 A 发生时, B 不发生。注意集合 B 可能比 A 大, 只减去 B 与 A 相同的子集元素。如 $A - B = \Phi$ 不能推出 $AB = \Phi$ 。

⑧独立: $P(AB) = P(A)P(B)$, 称为事件 A, B 独立, 特别注意独立概念是用概率来定义的。

其中事件 A 和 B 互斥 (不相容) 指的是 A 和 B 不能同时发生, 即 $AB = \Phi$; A 和 B 互逆 (对立) 也是指 A 和 B 不能同时发生, 即 $AB = \Phi$, 但是 A 和 B 必有一个发生, 且 $A \cup B = \Omega$ 。因此

A 和 B 互逆 $\Rightarrow A$ 和 B 互斥; A 和 B 互斥 $\nRightarrow A$ 和 B 互逆;

A 和 B 互逆 $\Leftrightarrow \bar{A}$ 和 $\bar{B}$ 互逆; A 和 B 互斥 $\nRightarrow \bar{A}$ 和 $\bar{B}$ 互斥。

三、事件的 8 大常用运算结论

$$\textcircled{1} A = A\Omega = A \cap \Omega = A(B \cup \bar{B}) = AB + A\bar{B}; AB \text{ 与 } A\bar{B} \text{ 不相容。}$$

$$\textcircled{2} A \cup B = A \cup (B - A)$$

$$\textcircled{3} A - B = A - AB = A(\Omega - B) = A\bar{B} \quad \text{称为差积转换公式, 如 } A \text{ 与 } B \text{ 不相容} \xrightarrow{AB=\Phi} A - B = A。$$

$$\textcircled{4} A + B = AB + A\bar{B} + \bar{A}B$$

$$\textcircled{5} A \cap (B - A) = \Phi$$

$$\textcircled{6} A \cup (B \cup C) = (A \cup B) \cup C$$

$$\textcircled{7} A \cup (B \cap C) = (A \cup B) \cap (A \cup C); \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$\textcircled{8} \overline{A \cup B} = \bar{A} \cap \bar{B}; \quad \overline{A \cap B} = \bar{A} \cup \bar{B}$$

对于第 8 个公式, 我们先在一个长方形内画出两个相交圆, 分别标注 1, 2, 3 部分, 再验证即可。

注意符号 $\cup \neq$ 数字 +, 例如: $A \cup B = A \cup (B - A) \neq A + (B - A) = B$; 符号 $\cap \neq$ 数字 $\times$ 。

例如: 任意数 $a, b, c \Rightarrow (a + b) - c = a + (b - c)$, 而对任意事件 A, B, C 一般没有类似的关系, 即 $(A + B) - C \neq A + (B - C)$; 又如 $A \cap (B - A) = \Phi \neq A \times (B - A) = AB - A$ 。所以, 读者应该把上述公式中的 + 与 $\times$ 号理解为并与交, 一般不遵循常规的四则运算规律。

第三节 事件发生的概率

一、概率的 4 种定义及其性质

1.1 概率的统计定义 (适应非等可能, 但试验次数无穷)

当 $n \rightarrow \infty$ 时, $f_n(A) \rightarrow$ 常数 $P(A)$, 则 $P(A)$ 称为事件 A 发生的概率。频率必须有多次试验才可计算, 而概率只要有一次试验即可计算。比如我在掷一次硬币之前就知道出现正面的概率为 $\frac{1}{2}$ 。我们投篮命中次数就是一个频率概念。

1.2 概率的古典定义 (古典概型或离散等可能概型)

试验结果一共有 n 个基本事件且每次试验各基本事件出现的可能性完全相同 (等可能性)。以 n 表示样本空间基本事件 (即集合中的元素总数) 的总数, m 表示事件 A 包含基本事件数, 则 $P(A) = \frac{m}{n}$, 即此时

概率等于频率，伯努利概型也属于这类概型。此类现象在客观情况中很普遍，研究的工具是利用排列与组合及加法与乘法原理。但对于主观因素有关的现象（非等可能性）就不适应，比如学生的考试成绩等等。

取球模型是指从 n 个球中任取 k 个球，每个球被取到的可能性相同，按照有放回和无放回以及考虑取球次序和不考虑取球次序，其基本事件总数如下：

取球方式		基本事件总数
有放回	考虑次序	n^k
	不考虑次序	C_{n+k-1}^k
无放回	考虑次序	P_n^k
	不考虑次序	C_n^k

以 3 个球任取两个为例分析如下：

■有放回，考虑次序，先把三个球编号为①②③

第一次取 ① 第二次分别取 ①, ②, ③ 共三种取法 ①①, ①②, ①③

第一次取 ② 第二次分别取 ①, ②, ③ 共三种取法 ②①, ②②, ②③

第一次取 ③ 第二次分别取 ①, ②, ③ 共三种取法 ③①, ③②, ③③

共有 9 种，等于 $3^2 = 9$ 。

■有放回，不考虑次序，先把三个球编号为①②③

则①②和②①、①③和③①、②③和③②为同一种取法，故共有 6 种，等于 $C_{n+k-1}^k = C_{3+2-1}^2 = 6$ 。

■无放回，考虑次序，先把三个球编号为①②③

第一次取 ① 第二次分别取 ②, ③ 共二种取法 ①②, ①③

第一次取 ② 第二次分别取 ①, ③ 共二种取法 ②①, ②③

第一次取 ③ 第二次分别取 ①, ② 共二种取法 ③①, ③②

共有 6 种，等于 $P_n^k = P_3^2 = 6$ 。

■无放回，不考虑次序，先把三个球编号为①②③

第一次取 ① 第二次分别取 ②, ③ 共二种取法 ①②, ①③

第一次取 ② 第二次分别取 ①, ③ 共二种取法 ②①, ②③

第一次取 ③ 第二次分别取 ①, ② 共二种取法 ③①, ③②

但①②、①③、②③为同一种取法，故 共有 3 种，等于 $C_n^k = C_3^2 = 3$ 。

放球模型是指把 n 个球放入 k 个盒中，按照每盒最多可以容纳一个球和可以容纳任意多个球以及球可分辨和不可分辨，其基本事件总数如下：

放球方式		基本事件总数
每盒最多可以容纳任意多个球	球可分辨	n^k
	球不可分辨	C_{n+k-1}^k
每盒最多可以容纳一个球	球可分辨	P_n^k
	球不可分辨	C_n^k

抽签原理 若 n 个签中有 m 个“有签”， $n-m$ 个“无签”， n 个人排队依次抽签（或一个人抽 n 次，每次抽一个），则第 k 个人（或某人第 k 次， $k=1,2,\dots,n$ ）抽到“有签”的概率都等于 $\frac{m}{n}$ 。由抽签原理知，其概率与各人抽签的先后次序无关，也与抽签放回还是不放回的方式无关，仅与“有签”所占比例有关。在考研数学中，抽签原理可以直接使用。

抽签原理的前提就是信息不公开, 抽签结果只有两个, 前面抽签后可以不放回, 也可以放回 (但放回的情形不是抽签原理的体现, 而是根据古典概型直接得出)。也就是前面抽的是白球还是红球 (或正品还是次品等等) 都不知道。所以任意一次抽的概率都相等。用全概率公式就可以证明。要是知道了前面的结果再抽就是条件概率了, 后面和前面不同。

1.3 概率的几何定义 (几何概型或连续等可能概型)

古典概型只适合有限离散等可能情形, 对于无限连续等可能情形, 则需要延伸它的定义, 即采用所谓的几何概型来描述: $P(A) = \frac{L(A)}{L(\Omega)}$, L 为相应几何测度, 可以为长度、面积或体积, 由此也可以看出概率是连续的, 它可以用天上掉馅饼来比喻, 我们用不同容积的器皿所接到的可能性是与器皿面积成正比的。几何概型是研考数学的一类常考题型, 研究的工具是利用微积分。注意, 等可能概型包括三种: 几何概型、古典概型和伯努利概型三种。

●几何概型的概率文氏图

在一个矩形区域 (表示概率等于 1) 内, 根据所给事件的概率关系画一系列圆, 并且把概率值写在圆域内, 利用面积的加减来分析概率问题。概率文氏图不能分析事件的集合关系, 只能分析事件的概率关系。比如: $P(AB) = 0 \nRightarrow AB = \Phi$; $P(B|A) = 0 \Rightarrow P(\overline{AB}) = 0 \nRightarrow A \subset B \text{ or } A \supset B$ 。但是, 已知事件关系却可以得到概率关系, 比如 $AB = \Phi \Rightarrow P(AB) = 0$ 。

概率文氏图一般需要掌握两种情形:

(a) 事件独立性情形。即两个圆的交集表示 $P(AB) = P(A)P(B)$ 。由此可以直观导出条件概率。

(b) 条件概率情形。如设 $P(A) = a$, $P(B) = b$, $P(AB) = c$, 则条件概率 $P(B|A) = \frac{c}{a}$ 正好对应事件

B 在事件 A 中的概率 c 与事件 A 的概率 a 之比; 同理, 则条件概率 $P(A|B) = \frac{c}{b}$ 则对应事件 A 在事件 B 的部分概率 c 与事件 B 的概率 b 之比。

1.4 概率的公理化定义 (普适定义, 1933 年)

设 E 是随机试验, Ω 为其样本空间, 以 E 中所有可能的随机事件组成的集合 F 为定义域, 对任意事件 A , 规定一个实值函数 $P(A)$, $A \in F$, 如果 $P(A)$ 满足下列三个条件, 就称 $P(A)$ 为事件 A 发生的概率。

- ① 非负性: $P(A) \geq 0$;
- ② 规范性: $P(\Omega) = 1$;
- ③ 可列可加性: 即: 设 $A_1, A_2, \dots, A_n$ 为两两互斥事件, 则

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)。$$

上述前三种定义都具有局域性, 是针对特定场合而言的, 只有公理化定义才具有一般性。公理化定义完全包含了前述三种定义的全部思想。公理化定义用实数描述事件, 则概率就是一种实变函数, 正是这种定义, 才使我们可以利用高等数学工具来研究概率问题。

6 种基本概率计算题型: 排列组合题型; 几何概率题型; 事件相关题型; 条件概率混合独立性题型;

全概率和逆概率题型; 伯努利题型。

二、条件概率

在事件 A 发生条件下, 事件 B 发生的概率, 称为条件概率, 用 $P(B|A)$ 表示。

以古典概型为例, 设样本空间 $\Omega = \{w_1, w_2, \dots, w_n\}$, 若事件 A 包含 m 个样本点, B 包含 s 个样本点, 而 s 个样本点中又有 k 个是属于 A 中的, 即事件 AB 包含 k 个样本点, 则

$$P(B|A) = \frac{AB \text{ 包含样本点的个数}}{A \text{ 包含样本点的个数}} = \frac{k}{m} = \frac{k/n}{m/n} = \frac{P(AB)}{P(A)} \Rightarrow P(B|A) = \frac{P(AB)}{P(A)}$$

$$P(A|B) = \frac{AB \text{ 包含样本点的个数}}{B \text{ 包含样本点的个数}} = \frac{k}{s} = \frac{k/n}{s/n} = \frac{P(AB)}{P(B)} \Rightarrow P(A|B) = \frac{P(AB)}{P(B)}$$

$P(B|A)$ 是在增加了条件 A 发生后, 在缩减的样本空间 Ω_A 中, B 的基本事件数与 Ω_A 之比。另外, 要特别注意条件概率中, 必须要求 $P(A) \neq 0$ 或 $P(B) \neq 0$, 即 $P(A) > 0$ 或 $P(B) > 0$ 。而在乘法公式 $P(AB) = P(A)P(B|A)$ 中就此要求。

三、概率的 5 类关系运算公式

利用概率文氏图理解。

3.1 加法公式

$$\textcircled{1} \quad \begin{aligned} &P(\Phi) = 0; \quad P(\Omega) = 1; \quad P(AB) \leq P(A); \quad P(A+B) \geq P(A) \\ &P(A) = 1 - P(\bar{A}) = P(AB) + P(A\bar{B}) \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad &P(A+B) = P(A) + P(B) - P(AB) = P(A-B) + P(B-A) + P(AB) \\ &= P(A\bar{B}) + P(B\bar{A}) + P(AB) = P(A) + P(\bar{A}B) = P(B) + P(A\bar{B}) = P(B) + P(A-B) \\ &= 1 - P(\overline{A+B}) = 1 - P(\bar{A} \cdot \bar{B}) \\ &P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \end{aligned}$$

$$\textcircled{3} \quad \text{若 } A_1, A_2, \dots, A_n \text{ 为两两互斥事件, 则 } P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

注意事件 $A_1, A_2, \dots, A_n$ 不一定独立。

3.2 减法公式

$$\begin{aligned} &P(A-B) = P(\overline{AB}) = P(A-AB) = P(A) - P(AB) \neq P(A) - P(B) \\ &\text{当 } A \supset B \Rightarrow P(A-B) = P(A) - P(B) \\ &AB = \Phi \Rightarrow P(A-B) = P(A) \\ &\text{当 } A, B \text{ 独立} \Rightarrow P(A-B) = P(\overline{AB}) = P(A)P(\bar{B}) = P(A)[1 - P(B)] \end{aligned}$$

3.3 乘法公式

$$\begin{aligned} &P(AB) = P(A)P(B|A) = P(B)P(A|B) \\ &P(ABC) = P(A)P(B|A)P(C|AB) \\ &P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}) \end{aligned}$$

3.4 条件概率公式

当条件“ $|B$ ”不变且 $P(B) \neq 0$ 时, 条件概率运算中 $P(A|B)$ 与 $P(A)$ 的全部公式完全相同。如:

$$\textcircled{1} \quad P(A|B) = 1 - P(\bar{A}|B)$$

$$\textcircled{2} \quad P[(A-C)|B] = P(A|B) - P(AC|B)$$

$$\textcircled{3} \quad P(\Omega|A) = 1 \quad P(A|A) = 1$$

$$\textcircled{4} \quad P(B|A) + P(\bar{B}|A) = 1$$

$$\textcircled{5} \quad P((B+C)|A) = P(B|A) + P(C|A) - P(BC|A)$$

$$\textcircled{6} \quad P(AB|C) = P(A|C)P(B|A|C) = P(A|C)P(B|AC)$$

如为互斥事件 (即: $P(BC) = 0$), 则 $P[(B+C)|A] = P(B|A) + P(C|A)$ 。

3.5 概率的单调与连续

① 单调性 $A \subset B \Rightarrow P(A) \leq P(B) \Rightarrow P(B-A) = P(B) - P(A)$ 。

② 概率是连续的, 且 $0 \leq P(A) \leq 1$ 。

四、全概率公式和逆概率公式 (贝叶斯公式)

4.1 一个划分

如果事件组 $A_1, A_2, \dots, A_n$ 满足 (1) $\bigcup_{i=1}^n A_i = \Omega$, (2) $A_i \cap A_j = \Phi$ ($i \neq j$), 称 $A_1, A_2, \dots, A_n$ (n 可以为无穷大) 为样本空间 Ω 的一个划分, 又称为 Ω 的一个完备事件组。正确理解一个划分是能否使用好全概率公式和贝叶斯公式的关键。注意, $A_1, A_2, \dots, A_n$ ($n > 2$) 为一个完备事件组, 则 $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ 就一定不是一个完备事件组。

4.2 全概率公式

已知原因求结果, 如果 $A_1, A_2, \dots, A_n$ 为一个划分, 并且 $A_1, A_2, \dots, A_n$ 是事件 B 发生的全部可能原因, 则

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i) = \sum_{i=1}^n P(A_i B)$$

简单地说, 求一般概率使用全概率公式。关键是要全面考虑第一阶段的所有可能结果构成一个划分。

4.3 逆概率公式 (贝叶斯公式)

已知结果求原因, 如果 $A_1, A_2, \dots, A_n$ 为一个划分, 并且 $A_1, A_2, \dots, A_n$ 是事件 B 发生的全部原因, 则

$$P(A_k|B) = \frac{P(A_k B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

简单地说, 求条件概率使用贝叶斯公式。关键是要全面考虑第一阶段的所有可能结果构成一个划分。

五、概率与事件的独立性的关系

5.1 两个事件独立

5.1.1 定义: $P(AB) = P(A)P(B|A) = P(A)P(B)$

● A, B 独立, 则 $P(\overline{A}\overline{B}) = P(\overline{A+B}) = 1 - P(A+B) = 1 - P(A) - P(B) + P(A)P(B)$
 $= [1 - P(A)] \cdot [1 - P(B)] = P(\overline{A})P(\overline{B}) \Rightarrow \overline{A}, \overline{B}$ 也独立。

● 任意事件 A 都与概率为 0 (不可能事件) 或 1 (必然事件) 的事件 B 相互独立。即

$$P(A\Phi) = P(A)P(\Phi) = 0; \quad P(A\Omega) = P(A)P(\Omega) = P(A); \quad P(\Phi\Omega) = P(\Phi)P(\Omega) = 0。$$

● 任意两个事件如有包含关系且它们的概率 $P(A) \neq 0, 1$, 则一定不相互独立。

● 概率为 1 或 0 的事件与任意一个事件相互独立。

5.1.2 两个事件独立的等价结论

设 $P(A) \neq 0, P(B) \neq 0$, 则两个事件 A, B 独立的等价结论有

$$(1) \quad P(B|A) = P(B|\overline{A}) = P(B) \quad \text{或} \quad P(A|B) = P(A|\overline{B}) = P(A)$$

$$(2) \quad P(B|A) + P(\overline{B}|\overline{A}) = 1$$

$$(3) \quad AB = \overline{A}\overline{B} \Rightarrow P(A|\overline{B}) + P(\overline{A}|B) = 2$$

$$(4) \quad A \text{ 与 } B \text{ 独立} \Leftrightarrow A \text{ 与 } \overline{B} \text{ 独立} \Leftrightarrow \overline{A} \text{ 与 } \overline{B} \text{ 独立。}$$

(5) 若 n 个事件相互独立, 则不含相同事件的事件组经并、差、交、逆运算后, 还是相互独立。

5.2 事件的互斥与独立的关系

A 与 B 互斥是指 A 与 B 不能同时发生, 即 $AB = \Phi$, 此时与事件的概率性质无关; 但是 A 与 B 独立是指 $P(AB) = P(A)P(B)$, 此时与事件的概率性质有关, 故“互斥”与“独立”并没有蕴含关系。

互斥的两个事件可以独立也可以不独立, 需要注意的是: 互斥的两个事件独立时, 则 $P(A)P(B) = 0$, 其余情形互斥的两个事件是不独立的。

5.3 多个事件 (事件组) 的独立

$$\textcircled{1} \quad \text{两两独立: } P(A_i A_j) = P(A_i)P(A_j) \quad (i \neq j)$$

$$\textcircled{2} \quad \text{相互独立: } P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}) \quad 1 \leq i_1 \leq i_2 \leq \cdots \leq i_k$$

如对三个事件 A, B, C

$$\text{两两独立: } P(AB) = P(A)P(B); \quad P(BC) = P(B)P(C); \quad P(AC) = P(A)P(C)$$

相互独立:

$$P(AB) = P(A)P(B); \quad P(BC) = P(B)P(C); \quad P(AC) = P(A)P(C); \quad P(ABC) = P(A)P(B)P(C)$$

可见, 相互独立, 则两两必独立, 反之不然。

A, B, C 相互独立, 则 $\overline{A}, \overline{B}, \overline{C}$ 也相互独立, 但他们的组合却不一定, 如 $\overline{AC}$ 与 $\overline{C}$ 就不独立, 因为

$$P(\overline{AC} \cdot \overline{C}) = P(\overline{AC} \cap \overline{C}) = P(\overline{AC \cup C}) = P(\overline{\Phi \cup C}) = P(\overline{C}) \neq P(\overline{AC})P(\overline{C}).$$

另外, 事件的独立性关系是指一般性关系, 而后面要讲述的相关关系仅仅指线性关系, 它只是一般关系中的一个特殊。另外, 特别注意: $[A, B \text{ 独立}] \neq [AB = \Phi]$, 这是因为独立是使用事件的概率来定义的, 而不是事件的相关关系。
概率为0或1的事件与任何事件均独立。

六、伯努利概型

6.1 单个伯努利试验

若一次试验 E 只有两个结果 A 或 $\overline{A}$ (如成功和失败), E 称为单个伯努利试验。

如 $P(A) = p$, 则 $P(\overline{A}) = 1 - p$ 。

6.2 n 重伯努利试验

当把 E 独立地重复进行 n 次, 称为 n 重伯努利试验。

n 重伯努利试验的具体条件有 5 个:

- (1) 独立, 指 $P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}) \quad 1 \leq i_1 \leq i_2 \leq \cdots \leq i_k$;
- (2) 重复, 指 $P(A_i) \equiv p$ 不变;
- (3) 要求等可能抽样;
- (4) 要求放回抽样, 与 $P(A) = p$ 是相容的;
- (5) 试验结果只有两种可能, 即对立结果。

形象记忆掌握法: n 重伯努利试验的特征是“独重回对等”。

如果不是抽样试验, 则 n 重伯努利试验的具体条件只有 4 个: “独重对等”。

6.3 伯努利概型公式 (二项分布 $X \sim B(n, p)$)

n 重伯努利试验中有许多事件发生, 其中事件 A 发生 (用 μ 表示) k 次的概率为

$$B(n, p) = P_n(k) = C_n^k p^k (1-p)^{n-k}. \text{ 注意: 当 } p=1 \text{ 时, } P_n(k) \equiv 1.$$

第二章 一维随机变量及其分布

一、随机变量

比如将一枚硬币抛三次，以 X 表示三次投掷中出现正面（用 H 表示正面， T 表示反面）的总次数，那么，对于样本空间 $\Omega = \{e\}$ 中的每一个样本点 e ， X 都有一个实数值与之对应，即

样本点 e	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
X 的值	3	2	2	2	1	1	1	0

这种对应并不要求反函数式的一一对应，但每一个数应能够一一对应同一事件，不能存在歧义，这样的 X 称为随机变量，其 3 个特征是：

第一，随机变量定义域为样本空间的全部基本事件；

第二，随机变量取值是随机的，它取每一个可能值有确定的概率（即分布函数）；

第三，随机变量是随机事件的人为数量化，而且这种数值只是一种符号表示。

二、随机变量的普适分布函数

2.1 分布函数定义诠释

一般情况下，人们只对某个区间内的概率感兴趣，即研究下列四种可能区间的概率

$$P\{x_1 < X \leq x_2\} \text{ 或 } P\{x_1 \leq X \leq x_2\} \text{ 或 } P\{x_1 \leq X < x_2\} \text{ 或 } P\{x_1 < X < x_2\}$$

读者只要利用一维坐标轴就能容易得出下列结论

$$\text{当 } \varepsilon \rightarrow 0 \Rightarrow \begin{cases} P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} \\ P\{x_1 \leq X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1 - \varepsilon\} \\ P\{x_1 \leq X < x_2\} = P\{X \leq x_2 - \varepsilon\} - P\{X \leq x_1 - \varepsilon\} \\ P\{x_1 < X < x_2\} = P\{X \leq x_2 - \varepsilon\} - P\{X \leq x_1\} \end{cases}$$

显然，我们只须定义一个 $P\{X \leq x\}$ 形式就可以了，其他区间形式都可以用它表示出来。

于是定义 $F(x) = P\{X \leq x\}$ 为 X 的普适分布函数。它就是 X 落在任意区间 $(-\infty, x]$ 上的概率，本质上是一个累积函数，对于离散点，采用叠加，对于连续点，使用一元积分。

引入随机变量的目的是从数量上来研究随机现象的统计规律，即把随机试验的不同结果用一个变量来表示，由于试验出现的结果是偶然的，因而随机变量的取值方式也是偶然的，试验前只能知道它的取值范围 $X \leq x$ ，试验后才能确定它的具体值 x 。另外，对于随机变量 X ，我们不仅要知道它取各种可能值的概率，更重要的是要知道 X 在任意区间 $[x_1, x_2]$ 内的取值分布规律，这正是分布函数所反映的内容——求事件的概率。

随机变量和分布函数共同架起了随机现象和高等数学之间的桥梁。

2.2 分布函数的 4 个重要性质

(a) $0 \leq F(x) \leq 1$ 。

(b) 单调不减。即 $x_2 > x_1 \Rightarrow F(x_2) \geq F(x_1)$ 。

(c) $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ 。

(d) 右连续。即对于任意实数 $x \Rightarrow F(x+0) = F(x)$ 。

实质上第一个性质已经包含在后 3 个性质中。上述 4 个性质是 $F(x)$ 是否为分布函数的充要条件。

作为分布函数的复合变量不能为单向取值, 因为这时会导致 $F(+\infty) = F(-\infty)$ 这一错误结论。比如 $F(x)$ 是 X 的分布函数, 则可能成为分布函数的是 $|F(-x)|$, $F(x^{2k+1}-1)$ 等形式, 但不能是 $F(x^{2k})$, $F(|x|)$ 等形式; 分布函数不能随便乘以常数, 如 $2F(x)$ 就不是分布函数。

2.3 分布函数要求右连续的数学根据

$$\left\{ \begin{array}{l} P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1) \\ P\{x_1 \leq X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1 - \varepsilon\} = F(x_2) - F(x_1 - 0) \\ P\{x_1 \leq X < x_2\} = P\{X \leq x_2 - \varepsilon\} - P\{X \leq x_1 - \varepsilon\} = F(x_2 - 0) - F(x_1 - 0) \\ P\{x_1 < X < x_2\} = P\{X \leq x_2 - \varepsilon\} - P\{X \leq x_1\} = F(x_2 - 0) - F(x_1) \\ P\{X < x_0\} = F(x_0 - 0) = \lim_{x \rightarrow x_0 - 0} F(x) \\ P\{X = x_0\} = P\{X \leq x_0\} - P\{X < x_0\} = F(x_0) - F(x_0 - 0) = F(x_0) - \lim_{x \rightarrow x_0 - 0} F(x) \\ P\{X > x_0\} = 1 - P\{X \leq x_0\} = 1 - F(x_0) \end{array} \right.$$

上述全部可能的表示中, 只有 $F(x-0)$ 形式, 但 $F(x-0) \neq F(x)$, 因为如 $F(x_1-0) = F(x_1)$, 那么, 当离散型在 x_1 点的概率不为零时, 等式 $P\{x_1 < X \leq x_2\} = P\{x_1 \leq X \leq x_2\}$ 就会出现矛盾, 故 $F(x)$ 不可能左连续。其中 $P\{X = x_0\} = F(x_0) - F(x_0 - 0) = F(x_0) - \lim_{x \rightarrow x_0 - 0} F(x)$ 是计算离散型分布函数的重要公式。

由于上式中根本不可能出现 $F(x+0)$ 的形式, $F(x+0) = F(x)$ 即 $F(x)$ 右连续在上述 5 种关系不存在, 故需要增加右连续规定, 即 $F(x_0+0) = F(x_0)$; $F(x_0-0) \neq F(x_0)$ 。当然, 由于连续型在一点的概率恒为零, 所以, 连续型分布函数左连续和右连续同时成立。正是要求 $F(x)$ 右连续, 才使 $F(x) = P\{X \leq x\}$ 成为分布函数的普适定义。所以, 在分布函数结论中, 等号一般和大于号放在一起, 以保证右连续, 请读者注意这个细节。

普适分布函数不仅可以描述离散型, 还可以描述连续型及其他非连续型。对连续型任一点的概率等于零, 而对非连续型任一点的概率不一定等于零。读者要重点掌握离散和连续两类随机变量的分布规律。任何随机变量都存在分布函数, 离散型的随机变量的分布函数一般为阶梯函数; 连续型随机变量的分布函数一定是连续函数。

不同的随机变量可以有相同的分布函数。比如抛一枚硬币, 令 $X_1 = \begin{cases} 1, & \text{出正面} \\ -1, & \text{出反面} \end{cases}$; $X_2 = \begin{cases} -1, & \text{出正面} \\ 1, & \text{出反面} \end{cases}$,

它们有共同的分布函数 $F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}, & -1 < x < 1, \text{ 且为离散型。} \\ 1, & x \geq 1 \end{cases}$

三、离散型随机变量的分布律

当随机变量所取的有限个或可列个值, 并能够按照由小到大的顺序排列时, 称为离散型随机变量。

当已知分布函数, 求分布律 (概率分布) 的计算方法如下

$$P\{X = x_i\} = F(x_i) - F(x_i - 0) = F(x_i) - \lim_{x \rightarrow x_i - 0} F(x).$$

设离散型随机变量 X 的可能取值为 $x_k (k=1, 2, \dots)$, 事件 $\{X = x_k\}$ 的概率为 $P\{X = x_k\} = p_k$, 离散型分布函数称为离散分布律, 一般用列表表示。注意 $\sum_{k=1}^{\infty} p_k = 1$ 。要求掌握的离散性分布律有 5 种: 0-1 分布, 伯努利二项分布, 泊松分布, 几何分布和超几何分布。

离散分布函数 $F(x) = P\{X \leq x\}$ 一般为阶梯函数。已知离散分布函数 $F(x)$, 根据分布函数的性质, 可以计算出离散分布律 $P\{X = x_k\}$; 反过来, 已知离散分布律 $P\{X = x_k\}$, 根据一维直角分割法(后述), 可以计算出离散分布函数 $F(x)$ 。

四、连续型随机变量的概率密度

4.1 连续型分布函数的性质

(a) 连续型 $F(x)$ 是连续函数 (左右均连续), 即 $F(x \pm 0) = F(x)$

(b) 连续型 $F(x)$ 几何意义是面积, 且 $F(x = x_0) \equiv 0$

(c) $F(+\infty) = \int_{-\infty}^{+\infty} f(t)dt = 1$ (必然事件), $F(-\infty) = \int_{-\infty}^{-\infty} f(t)dt = 0$ (不可能事件)

(d) $P\{a < X < b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = P\{a \leq X \leq b\} = F(b) - F(a) = \int_a^b f(t)dt$

(e) 要求掌握的连续型分布函数共有 3 种: 均匀分布, 指数分布和正态分布。

4.2 概率密度

$F(x) = \int_{-\infty}^x f(t)dt$ ($f(t) \geq 0$) 称为连续分布函数, $f(t) = F'(x)$ 称为概率密度, 或分布密度。

离散型分布函数反应在各个分布点上, 而连续型任意点上的分布函数值为 0, 显然不能反应其分布本质, 因而一般先求分布函数 $F(x)$ (即计算事件 $\{X \leq x\}$ 的概率), 然后对 $F(x)$ 求导得其相应的概率密度 $f(x)$ 来反应分布规律, 这一点和离散分布率是不同的。

连续型从分布函数求导而得密度时, 若 $F(x)$ 在 x 点可导, 则有 $f(x) = F'(x)$, 若 $F(x)$ 在 x 点不可导, 则直接令 $f(x) = 0$, 这不会影响分布函数的取值, 因此一般地密度函数与分布函数的求导关系为

$$f(x) = \begin{cases} F'(x), & F'(x) \text{ 存在} \\ 0, & F'(x) \text{ 不存在} \end{cases}. \text{ 故连续型的密度函数未必连续。}$$

连续型的密度函数不一定连续, 例如 $X \sim (a, b)$, 则 $f(x)$ 在 $x = a$ 或 b 两个端点处不连续, 所以, 一

般把均匀分布密度函数写成 $f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{other} \end{cases}$, 而不写成 $f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{other} \end{cases}$, 这一点读者要明确。

只有存在概率密度的随机变量才可能为连续型, 不能认为分布函数连续的随机变量就是连续型。也存在

既非离散也非连续的分布函数, 如 $F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$, 显然它既非阶梯函数也非连续函数。

但是即便密度函数存在也不一定是连续型,

如果 $g(x)$ 为连续函数, X 为连续型随机变量, $g(X)$ 不一定是连续型分布。如果 X 为离散型随机变量, 则 $g(X)$ 必是离散型分布, 离散型分布一般为阶梯函数。

『直角分割法』计算一维分布函数。『直角分割法』也适应二维分布

五、分布函数的重要结论

5.1 只有存在概率密度 (不恒为零) 的随机变量才称为连续型, 但不能错误认为分布函数连续的随机变量为连续型。如 $F(x) = \frac{1}{2}, x \geq 1$ 就不是连续型。

5.2 若 $F_1(x), F_2(x), \dots, F_n(x)$ 均是分布函数, 则

$$\sum_{i=1}^n a_i F_i(x) \left(a_i \geq 0, \sum_{i=1}^n a_i = 1 \right) \text{ 和 } \prod_{i=1}^n F_i(x) \text{ 仍然为分布函数。}$$

如果是连续型, 则上述求和 $\sum_{i=1}^n a_i = 1$ 以积分表出。

5.3 若 $f_1(x), f_2(x), \dots, f_n(x)$ 均是密度函数, 则

$$\sum_{i=1}^n a_i f_i(x) \left(a_i \geq 0, \sum_{i=1}^n a_i = 1 \right) \text{ 仍然为密度函数, 但 } \prod_{i=1}^n f_i(x) \text{ 不一定是密度函数。}$$

六、一维随机变量的 8 大分布

3 个离散分布 + 5 个连续分布。

1. 两点分布

又称 0-1 分布, 表示符号为 $B(1, p)$, 属于离散分布。

$$P(X=k) = C_1^k p^k (1-p)^{1-k} = p^k q^{k-1} \sim B(1, p), \quad k=0, 1.$$

模型: 试验变量 X 只有两种可能结果 (对立事件), 随机变量 X 使用 0 与 1 两种取值。如每次 A 发生的概率为 p , 共试验了 1 次, 求其中 A 发生的概率 (放回抽样)。

$$P(X=1) = p, \quad P(X=0) = 1-p = q$$

0-1 分布 $B\left(1, \frac{1}{2}\right)$ 的幂同分布, 比如 $X \sim B\left(1, \frac{1}{2}\right)$, 则 $X^2, X^3, \dots, X^n \sim B\left(1, \frac{1}{2}\right)$ 。

2. 伯努利二项分布

表示符号为 $B(n, p)$, 属于离散分布, 当 $n=1$, 为 0-1 分布。

$$P(X=k) = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k} \sim B(n, p), \quad k=0,1,2,\dots,n$$

模 型: 随机试验结果只有两种, 如每次 A 发生的概率为 p , 共试验了 n 次独立试验, 求其中 A 发生 k 次的概率 (放回抽样)。

设 $X \sim B(n, p)$, 则使得概率 $P(X=k)$ 达到最大的 k 为

$$k = \begin{cases} (n+1)p \text{ or } (n+1)p-1, & (n+1)p \text{ 为正整数} \\ \lfloor (n+1)p \rfloor, & (n+1)p \text{ 为非整数} \end{cases}$$

若 $X \sim B(n, p)$, 则 $Y = n - X \sim B(n, 1-p)$

3. 泊松分布

表示符号为 $P(\lambda)$, 属于离散分布。

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \sim P(\lambda), \quad k=0,1,2,\dots$$

模 型: 满足下列条件的随机质点流 (一串重复出现的事件) 称为泊松流。

(1) 在时间 $(t, t+\Delta t)$ 内流过质点数的概率仅与 Δt 有关, 而与 t 无关;

(2) 不相交的时间间隔内流过的质点数彼此独立;

(3) 在充分短的一瞬间只能流过一个或没有质点流过, 要流过 2 个或 2 个以上质点几乎是不可能的。可以证明泊松流在单位时间内流过质点数便服从泊松分布。

例如: 单位时间内放射性物质放射出的粒子数; 单位时间内某电话交换台接到的呼唤次数; 单位时间内走进商店的顾客数等等, 均可认为它们服从泊松分布。

当 p 很小时, 有 $P(\lambda) = \lim_{n \rightarrow \infty} B(n, p)$, 其中 $\lambda = np$, 即泊松分布是伯努利二项分布的极限形式。

设 $X \sim P(\lambda)$, 则使得概率 $P(X=k)$ 达到最大的 k 为: $k = \begin{cases} \lambda \text{ or } \lambda-1, & \lambda \text{ 为正整数} \\ \lfloor \lambda \rfloor, & \lambda \text{ 为非整数} \end{cases}$ 。

4. 几何分布

表示符号为 $G(p)$, 属于离散分布。

$$P(X=k) = p(1-p)^{k-1} = pq^{k-1} \sim G(p), \quad k=1,2,3,\dots$$

模 型: 随机试验结果只有两种, 如每次 A 发生的概率为 p , 试验一直继续, 直到 A 发生为止, 求第 k 次 (放回抽样) A 才发生的概率。

5. 超几何分布

表示符号为 $H(n, M, N)$, 属于离散分布。

$$P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \sim H(n, M, N), \quad k=0,1,2,\dots,\min(M,n)$$

模 型: N 个元素分为 N_1 和 N_2 两类, 从中取 n 件 (不放回, 如放回抽样则是二项分布模型), 其中含有 k 个

第一类元素的概率为 $\frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n} (k=0,1,2,\dots)$ 。

当 N 很大, 而 n 相对 N 较小, 设 $p = \frac{M}{N}$, 则有

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k} \Leftrightarrow H(n, M, N) \approx B(n, p)$$

6. 均匀分布

表示符号为 $U(a, b)$, 属于连续分布。

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{other} \end{cases} \sim U(a, b) \quad \text{注意区间为开区间, 端点的分布密度值取零}$$

$$F(x) = \int_{-\infty}^x f(x) dx = \begin{cases} 0, & x < a \\ \int_a^x f(x) dx = \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

$$P\{a \leq x_1 < X \leq x_2 \leq b\} = F(x_2) - F(x_1) = \frac{x_2 - x_1}{b - a}$$

模型: 设随即变量 X 的值落在 (a, b) 内, 其内取值具有“等可能”性。

7. 指数分布

表示符号为 $E(\lambda)$, 属于连续分布。

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases} \sim E(\lambda) \quad F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

模 型: 在实践中, 如果随机变量 X 表示某一随机事件发生所需等待的时间, 则一般 $X \sim E(\lambda)$ 。例如, 某电子元件直到损坏所需的时间 (即寿命); 随机服务系统中的服务时间; 在某邮局等候服务的等候时间等等均可认为是服从指数分布。

指数分布计算中常用到 Γ 函数公式 $\Gamma(n+1) = \int_0^{+\infty} x^n e^{-x} dx = n!$ 。

指数分布的无记忆性定理: $P(X > s+t | X > s) = P(X > t)$

8. 正态分布

8.1 正态分布的密度和分布函数

表示符号为 $N(\mu, \sigma^2)$, 属于连续分布。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \sim N(\mu, \sigma^2), \quad x \in (-\infty, +\infty) \Rightarrow Y = aX + b \sim N[a\mu + b, (a\sigma)^2]$$

● 模 型：在实践中，如果随机变量 X 表示许许多多均匀微小随机因素的总效应，则它通常将近似地服从正态分布。如：测量产生的误差；弹着点的位置；噪声电压；产品的尺寸等等均可认为近似地服从正态分布。尽管它来源于连续型，但它是任何分布在样本数一般大于 45 时的极限分布。而且，根据中心极限定理，若干个未知分布的随机变量之和近似地服从正态分布，它是数理统计的基础，是概率与数理统计中的第一大分布。

● 当 $\mu = 0, \sigma = 1 \Rightarrow f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sim N(0, 1)$ ，称为标准正态分布。此时分布函数及其性质为

$$\Phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \Rightarrow \begin{cases} \Phi(-x) = 1 - \Phi(x) \\ \Phi(0) = \frac{1}{2} = P\{X > 0\} = P\{X \leq 0\} \\ \Phi\left(\frac{X - \mu}{\sigma}\right) \sim N(0, 1) \\ P(x_1 < X \leq x_2) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right) \end{cases}$$

评 注 8 大分布产生的背景如下，伯努利试验产生的分布有： $0-1$ 分布， $B(n, p)$ ， $G(p)$ ， $H(n, M, N)$ ；泊松流产生的分布有： $P(\lambda)$ ， $E(\lambda)$ ；误差产生的分布有： $U(a, b)$ ， $N(\mu, \sigma^2)$ 。

8.2、数理统计中的 4 个分位数

$$(1) \text{ 上分位数 } P(X > z_\alpha) = \alpha \Rightarrow \int_{z_\alpha}^{+\infty} N(\mu, \sigma^2) dx = \alpha$$

$$(2) \text{ 下分位数 } P(X \leq z_\alpha) = \alpha \Rightarrow \int_{-\infty}^{z_\alpha} N(\mu, \sigma^2) dx = \alpha$$

评 注 无论哪种分位数，对标准正态分布都有： $z_{1-\alpha} = -z_\alpha$

切记标准正态分布的查表中使用的 $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = P(Z \leq z)$ 是下分位数；其他三种抽样分

布的查表中则使用的是上分位数，即
$$\begin{cases} P\{\chi^2(n) > \chi_\alpha^2(n)\} = \alpha \\ P\{t(n) > t_\alpha(n)\} = \alpha \\ P\{F(n_1, n_2) > F_\alpha(n_1, n_2)\} = \alpha \end{cases}。$$

标准化公式： $\frac{X - \mu}{\sigma} \sim N(0, 1)$

七、一维随机变量函数 $Y = f(X)$ 的分布的求法

如果 $g(x)$ 为连续函数， X 为连续型随机变量， $Y = g(X)$ 不一定是连续型分布，所以题中必须明显给出或说明随机变量存在概率密度，以表征 $Y = g(X)$ 为连续型函数分布。如果 X 为离散型随机变量，则

$g(X)$ 必是离散型分布, 离散型分布一般为阶梯函数。

一、离散型随机变量的函数分布的计算方法

已知 X 的分布律 $P\{X = x_k\} = p_k (k=1, 2, \dots)$, 则 X 的函数 $Y = g(X)$ 的分布律按以下原则确定:

1. 等价原则

若 $y_k = g(x_k) (k=1, 2, \dots)$ 互不相同, 则 $P\{Y = y_k\} = P\{X = x_k\} = p_k (k=1, 2, \dots)$ 。

2. 相加原则

若 $y_k = g(x_k) (k=1, 2, \dots)$ 部分值相同, 则相同的值对应的概率相加, 而不同的值仍按照“等价原则”。

二、连续型随机变量的函数分布的计算方法

2.1 分布函数微分法

分布函数微分法是求 $Y = g(X)$ 分布的一般的解法

(1) 首先写出 $F_X(x)$ 并确定 Y 的值域 $[a, b]$, 也可以是开区间或半开半闭区间。

(2) $y < a \Rightarrow F_Y(y) = 0; \quad y \geq b \Rightarrow F_Y(y) = 1;$

(3) $a \leq y < b$, 根据分布函数定义求, 即

$$F(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = \int_{g(X) \leq y} f(x) dx \Rightarrow f(y) = [F(y)]'.$$

2.2 公式法

当 $y = g(x)$ 单调时, 由分布函数微分法可推导得到公式

$$f_Y(y) = \begin{cases} f_X[h(y)]|h'(y)|, & y \in (\alpha, \beta) \\ 0, & \text{other} \end{cases}$$

其中 $h(y)$ 为 $y = g(x)$ 的反函数, $\alpha = \min\{g(-\infty), g(+\infty)\}$, $\beta = \max\{g(-\infty), g(+\infty)\}$ 。

注意推广的 $Y = g(X)$ 的公式法: 如果 $g(x)$ 逐段严格单调, 则 $f_Y(y) = \sum f_X[h_i(y)]|h'_i(y)|$ 。

若 $f_X(x)$ 在有限区间 $[a, b]$ 上大于零, 而在其它点处皆为零, 则

$$\alpha = \min\{g(a), g(b)\}; \quad \beta = \max\{g(a), g(b)\}$$

若 $y = g(x)$ 不为单调函数, 则划分为单调区间, 在每个单调区间使用上述公式, 然后将结果合并。

如果 X 为连续型, 则 $Y = aX + b$ 也是连续型, 根据上述公式容易得出: $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$ (由

函数分布的导数公式也可直接导出), 如果 X 为离散型, 则 $Y = aX + b$ 却不一定为同一类型的离散型, 如 X 服从泊松分布, $Y = aX + b$ 就不再是泊松分布。

2.3 智轩积分转换法

设随机变量 X 的概率密度为 $f_X(x)$, $g(x)$ 为分段连续或分段单调函数, $Y = g(X)$, 对任何非负连续

函数 $h(x)$, $\int_{-\infty}^{+\infty} h[g(x)]f(x)dx = \sum_{i=1}^n \int_{\alpha_{i-1}}^{\alpha_i} h(y)p_i(y)dy$ 成立, 则随机变量 Y 的密度函数 $f_Y(y)$

$$f_Y(y) = \begin{cases} p_1(y), & \alpha_0 < y \leq \alpha_1 \\ p_2(y), & \alpha_1 < y \leq \alpha_2 \\ \dots & \\ p_m(y), & \alpha_{m-1} < y \leq \alpha_m \\ 0, & \text{other} \end{cases}, \quad \text{它的理论根据来自数学期望的思想。}$$

三、两种分布的组合

一般是一个离散和一个连续, 使用全概率公式。

第三章 多维随机变量及其分布

一、二维随机变量(向量)的分布函数的普适定义

定义 $F(x, y) = P\{X \leq x, Y \leq y\} \xrightarrow["Y \leq y^* = B"]{"X \leq x^* = A"} = P(AB)$ 为二维随机变量 (X, Y) 的分布函数, 又称联合分布函数。

联合分布函数的性质

- (1) $0 \leq F(x, y) \leq 1$;
- (2) $F(x, y)$ 对 x 和 y 都是单调非减的, 如 $x_1 > x_2 \Rightarrow F(x_1, y) \geq F(x_2, y)$;
- (3) $F(x, y)$ 对 x 和 y 都是右连续;
- (4) $F(+\infty, +\infty) = \lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1, F(-\infty, -\infty) = F(x, -\infty) = F(-\infty, y) = 0$

联合分布函数的几何意义表示 $F(x, y)$ 在 (x, y) 的函数值就是随机点 (X, Y) 在 $X = x$ 左侧和 $Y = y$ 下方的无穷矩形内的概率。

对有限矩形域有 $P\{x_1 < X < x_2, y_1 < Y < y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$

二、2 维离散型随机变量的分布函数

二维离散型随机变量的分布函数使用联合分布律、边缘分布律与条件分布律, 简称 **3 律** 来描述。

2.1 联合分布律

设 (X, Y) 的一切可能取值为 (x_i, y_j) ; $i, j = 1, 2, \dots$, 且 (X, Y) 取各对可能值的概率为

$P\{X = x_i, Y = y_j\} = p_{ij}$, 离散型 $F(x, y)$ 的结果形式为一系列的点的值, 一般使用表格表示。

2.2 边缘分布律

设事件 $A_i = \{X = x_i\}$, $B_j = \{Y = y_j\}$, 根据全概率公式有

$P\{X = x_i\} = P(A_i) = \sum_{j=1}^n P(B_j)P(A_i | B_j) = \sum_{j=1}^n P(A_i B_j) = \sum_{j=1}^n p_{ij} = p_{i\cdot}$, 称为 X 方向边缘分布律。

$P\{Y = y_j\} = P(B_j) = \sum_{i=1}^n P(A_i)P(B_j | A_i) = \sum_{i=1}^n P(A_i B_j) = \sum_{i=1}^n p_{ij} = p_{\cdot j}$, 称为 Y 方向边缘分布律。

2.3 条件分布律

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

【智轩二维直角分割法】

在某局部区域中, 已知 2 维随机变量不为零的分布密度 $f(x, y)$ (这个局部区域又称为正概率点区域), 求全部区域的分布函数 $F(x, y)$ 问题是一个难点。作者创立的直角分割法可以方便清晰地解决这类题型。二维直角分割法秘诀如下,

- (1) 如果正概率点区域在 x 和 y 两个方向都有界, 则需要将全平面区域划分为 5 类积分区域, 在每类区域中求 $F(x, y)$ 时, 积分区域为直角分割区域和正概率点区域的交集。

- 第 1 类积分区域 点 (x, y) 的直角分割区域与 $f(x, y)$ 正概率区域无交集, $F(x, y) = 0$;
- 第 2 类积分区域 点 (x, y) 的直角分割区域画在与 $f(x, y)$ 正概率点区域 y 方向的外边, 积分区域为直角分割区域和正概率区域的交集部分, 显然这时相当于求 X 的边缘分布函数 $F(x, y) = F_X(x)$;
- 第 3 类积分区域 点 (x, y) 的直角分割区域画在与 $f(x, y)$ 正概率点区域 x 方向的外边, 积分区域为直角分割区域和正概率区域的交集部分, 显然这时相当于求 Y 的边缘分布函数 $F(x, y) = F_Y(y)$;
- 第 4 类积分区域 点 (x, y) 的直角分割区域画在 $f(x, y)$ 正概率点区域的内部, 积分区域为直角分割区域和正概率区域的交集部分;
- 第 5 类积分区域 点 (x, y) 的直角分割区域包含整个正概率点区域的全部, 积分区域是正概率点区域本身, 显然此时有 $F(x, y) = 1$ 。

(2) 如果正概率点区域在 x 和 y 两个方向有一个区间无界, 由于直角分割区域顶点无法画在该无界区间的外部, 则只需将全部区域划分为 3 类积分区域, 即没有第 2 类和第 5 类, 或者没有第 3 类和第 5 类。在每类区域中 $F(x, y)$ 积分区域仍为直角分割区域和正概率点区域的交集。

三、二维连续型随机变量的分布函数

二维连续型随机变量的分布函数使用联合概率密度、边缘概率密度与条件概率密度, 简称 3 密来描述。二维连续型的两个分量还是连续型随机变量, 反正不成立。

3.1 联合分布函数与联合概率密度

连续型联合分布函数:

$$F(x, y) = P\{X \leq x, Y \leq y\} = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = \iint_{\{\text{直角分割区域}\} \cap \{\text{正概率区域} D\}} f(u, v) du dv$$

区域 D 按照智轩直角分割法确定。且有联合概率密度:

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

3.2 边缘概率密度

$$\begin{aligned} F_X(x) = F(x, +\infty) &= \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f(x, y) dy \right] dx \Rightarrow f_X(x) = \frac{dF_X(x)}{dx} = \int_{-\infty}^{+\infty} f(x, y) dy \\ F_Y(y) = F(+\infty, y) &= \int_{-\infty}^y \left[\int_{-\infty}^{+\infty} f(x, y) dx \right] dy \Rightarrow f_Y(y) = \frac{dF_Y(y)}{dy} = \int_{-\infty}^{+\infty} f(x, y) dx \end{aligned}$$

评注 二维连续型 (X, Y) 的两个分量 X, Y 还是连续型, 但两个分量都是连续型的随机变量构成的二维随机变量却不一定是连续型, 即可能成为既非连续型, 又非离散型。

3.3 条件概率密度

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}; \quad f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

3.4 连续型分布的概率密度、边缘密度和条件密度函数的关系

- 乘法公式
$$f(x, y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y)。$$
 $(P_{ij} = P_{i.} P_{.j|i} = P_{.j} P_{i|j})$

● 全概率公式

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\infty}^{+\infty} f_Y(y) f_{X|Y}(x|y) dy$$

● 贝叶斯公式

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

● 二维随机变量 (X, Y) 的联合分布唯一地确定两个边缘分布、条件分布；但反之不然。若 X, Y 独立，由两个边缘分布可以确定联合分布；若 X, Y 不独立，则由一个边缘分布再加上一个对应的条件分布才能确定联合分布（参看上述乘法公式）。

四、2 大二维连续型分布函数

4.1 二维均匀分布

$$f(x, y) = \begin{cases} \frac{1}{S_D}, & (x, y) \in D \\ 0, & \text{other} \end{cases} \sim U(D)$$

如果 (X, Y) 服从矩形区域 $D = \{(x, y) | a \leq x \leq b, c \leq y \leq d\}$ 上的均匀分布，则两个边缘分布和两个条件分布都是对应的一维均匀分布，而且 X, Y 独立。反之， X, Y 均服从一维均匀分布，如相互独立则 (X, Y) 服从二维均匀分布，如不独立，则 (X, Y) 不一定服从二维均匀分布。

若 (X, Y) 服从非矩形区域、圆形区域 $D = \{(x, y) | x^2 + y^2 \leq r^2\}$ 等等上的均匀分布，则两个边缘分布都不是均匀分布，但两个条件分布都是均匀分布。

4.2 二维正态分布

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right]} \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$$

二维正态分布的两个边缘分布和两个条件分布及其两个分量的线性组合仍然是正态分布。

两个边缘分布为：

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \sim N(\mu_1, \sigma_1^2) \quad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \sim N(\mu_2, \sigma_2^2)$$

两个条件分布为：

$$(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho) \Rightarrow \begin{cases} f_{X|Y} \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2), \sigma_1^2(1 - \rho^2)\right) \\ f_{Y|X} \sim N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2)\right) \end{cases}$$

从上式可以看出，当 $\rho = 0$ ，即 X, Y 独立或不相关时，两个正态边缘分布和条件分布相同。

两个分量 X, Y 的任意线性组合为：

$$Z = aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$$

是一维正态分布；但任意两个单一正态随机变量的线性组合却不一定是正态分布；两个边缘分布都是正态分布的二维随机变量也不一定是正态分布，只有在相互独立的情形下才满足正态分布形式。

如果 X_1, X_2 为来自总体 $X \sim N(0, \sigma^2)$ 的简单随机变量, 则 $X_1 + X_2$ 和 $X_1 - X_2$ 必定相互独立。

n 维正态分布具有下列四种重要性质:

(1) n 维正态变量 $(X_1, X_2, \dots, X_n)$ 的每一个分量都是一维正态分布, 但 $X_1, X_2, \dots, X_n$ 不一定相互独立。反之, $X_1, X_2, \dots, X_n$ 都是一维正态分布, 则只有相互独立条件下, $(X_1, X_2, \dots, X_n)$ 才能构成 n 维正态分布。

(2) n 维随机变量 $(X_1, X_2, \dots, X_n)$ 服从 n 维正态分布的充要条件是 $c_1 X_1 + c_2 X_2 + \dots + c_n X_n$ 服从一维正态分布。 $X_1, X_2, \dots, X_n$ 都是一维正态分布, 只有相互独立条件下, $c_1 X_1 + c_2 X_2 + \dots + c_n X_n$ 服从一维正态分布, 且具有可加性。

(3) $(X_1, X_2, \dots, X_n)$ 是 n 维正态分布, 则 $Y_1, Y_2, \dots, Y_n$ 是 $X_1, X_2, \dots, X_n$ 的任意线性组合, 则 $(Y_1, Y_2, \dots, Y_n)$ 也是 n 维正态分布。

(4) $(X_1, X_2, \dots, X_n)$ 是 n 维正态分布, 则 $X_1, X_2, \dots, X_n$ 相互独立与两两不相关等价。但是即使 $\rho_{XY} = 0$ 也不能判断 X, Y 相互独立, 因为 (X, Y) 未必是正态分布。

4.3 连续型分布的概率密度、边缘密度和条件密度函数的关系

● 乘法公式

$$f(x, y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y). \quad (P_{ij} = P_{i.} P_{.j| i} = P_{.j} P_{i| j})$$

● 全概率公式

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\infty}^{+\infty} f_Y(y) f_{X|Y}(x|y) dy$$

● 贝叶斯公式

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

● 二维随机变量 (X, Y) 的联合分布唯一地确定两个边缘分布、条件分布; 但反之不然。若 X, Y 独立, 由两个边缘分布可以确定联合分布; 若 X, Y 不独立, 则由一个边缘分布再加上一个对应的条件分布才能确定联合分布 (参看上述乘法公式)。

五、随机变量的独立性与 3 分 3 律 3 密

二元分布有离散型的联合、边缘和条件分布律 3 种形式和连续型的联合、边缘和条件 3 种密度函数, 简称 3 分 3 密, 它们的一种重要关系就是独立性。

5.1 独立性的一般定义形式

任意 $x_1, \dots, x_n$ 的联合分布函数 $F(x_1, \dots, x_n)$ 满足

$$F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n) \text{ 时, 称 } x_1, \dots, x_n \text{ 相互独立。}$$

注意, 可以证明, 这个定义与前面的用事件的概率来定义事件之间的独立性是完全等价的。

★ 二维离散型: X, Y 相互独立的充要条件是 $p_{ij} = p_{i.} p_{.j}$

★ 二维连续型: X, Y 相互独立的充要条件是 $f(x, y) = f_X(x) f_Y(y)$

5.2 规则区域上的独立性

如果 $f(x, y)$ 在规则区域, 如矩形区间等, 具有分离变量形式, 即

$$f(x, y) = g(x)h(y) \quad [x \in (a, b), y \in (c, d)], \text{ 则 } X, Y \text{ 一定相互独立。}$$

(a) 如 $f(x, y) = 8xy \quad (0 \leq x \leq 1, 0 \leq y \leq 1)$ 中 X, Y 就一定独立。

(b) 如 $f(x, y) = 8xy \quad (0 \leq x \leq y, 0 \leq y \leq 1)$, 存在不规则区间, 故 X, Y 不独立。

若上述两个条件 规则区域或分离变量形式 一个都不满足, 则一般不独立。注意 $g(x), h(y)$ 不是边缘分布。

5.3 二维正态型随机变量相互独立的充要条件

二维正态型随机变量 (X, Y) 相互独立的充要条件是相关系数 $\rho = 0$, 即 X, Y 不相关。

如果 $X_i \sim N(\mu_i, \sigma_i^2)$, 且 $\{X_i\}$ 相互独立, 则 $Z = \sum_{i=1}^n k_i X_i \sim N\left(\sum_{i=1}^n k_i \mu_i, \sum_{i=1}^n (k_i \sigma_i)^2\right)$

5.4 随机变量函数的相互独立性

(1) 设随机向量 $(X_1, \dots, X_m)$ 和 $(Y_1, \dots, Y_n)$ 及 $(X_1, \dots, X_m; Y_1, \dots, Y_n)$ 满足下式

$$F(x_1, \dots, x_m; y_1, \dots, y_n) = F_1(x_1, \dots, x_m)F_2(y_1, \dots, y_n)$$

则称 $(X_1, \dots, X_m)$ 与 $(Y_1, \dots, Y_n)$ 相互独立; 此时, X_i 与 Y_i 必相互独立; 并且, 任意函数分布

$g(X_1, \dots, X_m)$ 与 $h(Y_1, \dots, Y_n)$ 也相互独立,

(2) 如随机变量 X_1, X_2 相互独立, 则随机变量的函数 $f(X_1)$ 与 $g(X_2)$ 必相互独立, 但 $f(X_1, X_2)$

与 $g(X_1, X_2)$ 却不一定独立。

(3) $X \sim N(0, 1), Y \sim \chi^2(m), Z \sim \chi^2(n) (m \neq n)$, 则 X, Y, Z 相互独立。

5.5 四类可加性分布 (二泊正卡)

(1) X, Y 相互独立, $X \sim B(m, p), Y \sim B(n, p)$, 则 $Z = X + Y \sim B(m+n, p)$

(2) X, Y 相互独立, $X \sim P(\lambda_1), Y \sim P(\lambda_2)$, 则 $Z = X + Y \sim P(\lambda_1 + \lambda_2)$

但泊松分布不存在线性关系, 即 $Y = aX + b$ 不再是泊松分布。

(3) X, Y 独立, $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, 则

$$Z = aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

如果 (X, Y) 联合分布是正态, X, Y 不独立, 则

$$Z = aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$$

(4) X, Y 相互独立, $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$, 则 $Z = X + Y \sim \chi^2(n_1 + n_2)$

一般来说, 如果 (X, Y) 的联合分布律中有某一项为零, 而该项所在的行和列的其余元素不全为零, 则 (X, Y) 必然不独立。

事件 A, B 独立等价于它们相应的二值变量 X, Y 独立。

五、4 大二维随机变量的函数的分布及其模型

■ 函数分布的一般求法

一、离散型条形表和倒表法

对于离散型, 函数分布律一般使用条形表和倒表法求子, 对于连续型函数分布的密度一般使用雅可比变换法和智轩积分变化法。

二、连续型雅可比变换法

一般地, 设 (X, Y) 的概率密度为 $f(x, y)$, 如果 $U = g(X, Y), V = \varphi(X, Y)$ 存在唯一的反函数:

$X = X(U, V), Y = Y(U, V)$, 利用雅可比微元变换 $J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$, 可得变换后的分布函数形式为

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

比如设 $Z = X + Y$, 为应用函数的变换公式, 现人为添加恒等式 $X = X$, 则构成

$$\begin{cases} x = x \\ z = x + y \end{cases} \Rightarrow \begin{cases} x = x \\ y = z - x \end{cases} \Rightarrow J = \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

从而 (X, Z) 的概率密度为 $f(x, z-x)|J| = f(x, z-x)$, $f_z(z)$ 可以看成是 (X, Z) 的边缘分布, 故

$$f_z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx.$$

也可以这样考虑, 现人为添加恒等式 $Y = Y$, 则构成

$$\begin{cases} y = y \\ z = x + y \end{cases} \Rightarrow \begin{cases} y = y \\ x = z - y \end{cases} \Rightarrow J = \begin{vmatrix} \frac{\partial y}{\partial y} & \frac{\partial y}{\partial z} \\ \frac{\partial x}{\partial y} & \frac{\partial x}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

从而 (Y, Z) 的概率密度为 $f(z-y, y)|J| = f(z-y, y)$, $f_z(z)$ 可以看成是 (Y, Z) 的边缘分布, 故

$$f_z(z) = \int_{-\infty}^{+\infty} f(z-y, y) dy.$$

利用类似手段, 设 $a > 0, b > 0$, 可以得到 $Z = aX + bY$; $Z = aXY$; $Z = \frac{aX}{bY}$ (or $\frac{bY}{aX}$) 的密度函数。

三、【连续型智轩积分变换法】

设二维随机变量 (X, Y) 的联合密度函数为 $f(x, y)$, 则 $Z = g(X, Y)$ 的一维密度函数 $f_z(z)$ 有与一元

随机变量的类似的积分变换法。设 $g(x, y)$ 是连续的实函数, 对任意非负连续函数 $h(z)$, 下式成立:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h[g(x, y)] f(x, y) dx dy = \sum_{i=1}^m \int_{\alpha_{i-1}}^{\alpha_i} h(z) p_i(z) dz$$

$$\text{则 } Z = g(X, Y) \sim f_Z(z) = \begin{cases} p_1(z), & \alpha_0 < z \leq \alpha_1 \\ p_2(z), & \alpha_1 < z \leq \alpha_2 \\ \dots \\ p_m(z), & \alpha_{m-1} < z \leq \alpha_m \end{cases}.$$

■ 四类二维函数分布

一、备用模型

备用模型事件表示为 $Z = aX + bY$, 分布函数 $F_Z(z) = P(Z \leq z) = P(aX + bY \leq z)$ 。

1.1 离散型备用模型

一般方法: 代入 X, Y 的各种可能值, 再计算对应每一可能值的概率和。

1.2 连续型备用模型

使用基本定义推导如下, 设 (X, Y) 的概率密度为 $f(x, y)$, 则连续型 $Z = aX + bY$ 的分布函数定义为

$$F_Z(z) = P\{Z \leq z\} = \iint_{ax+by \leq z} f(x, y) dx dy.$$

其中 $ax + by \leq z$ 是 $ax + by = z$ 及其左下方的半平面, 则

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} = P\{aX + bY \leq z\} = \iint_{ax+by \leq z} f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} dy \int_{-\infty}^{\frac{z-by}{a}} f(x, y) dx \xrightarrow[\frac{dx}{a} = \frac{1}{a} du]{ax+by=u} = \frac{1}{|a|} \int_{-\infty}^{+\infty} dy \int_{-\infty}^z f\left(\frac{u-by}{a}, y\right) du = \int_{-\infty}^z \left[\frac{1}{|a|} \int_{-\infty}^{+\infty} f\left(\frac{u-by}{a}, y\right) dy \right] du \\ &\Rightarrow f_Z(z) = F'_Z(z) = \frac{1}{|a|} \int_{-\infty}^{+\infty} f\left(\frac{z-by}{a}, y\right) dy \quad (3y \text{ 型}) \end{aligned}$$

希望读者仿照上述方法反复推导 2 次, 领会其一般思想, 切不可硬背。如果存在非正规区域 (即: 积分不能用一项表示出来), 则需要使用平移法划分为若干个正规区域, 后使用智轩平移法。

$$\text{由于 } X, Y \text{ 的轮换对称性, 易知 } f_Z(z) = F'_Z(z) = \frac{1}{|b|} \int_{-\infty}^{+\infty} f\left(x, \frac{z-ax}{b}\right) dx \quad (3x \text{ 型})$$

$$\text{即得连续型备用模型公式 } f_Z(z) = \frac{1}{|a|} \int_{-\infty}^{+\infty} f\left(\frac{z-by}{a}, y\right) dy = \frac{1}{|b|} \int_{-\infty}^{+\infty} f\left(x, \frac{z-ax}{b}\right) dx$$

如果 X, Y 独立, 则

$$Z = aX + bY \Rightarrow f_Z(z) = \frac{1}{|b|} \int_{-\infty}^{+\infty} f_X(x) f_Y\left(\frac{z-ax}{b}\right) dx = \frac{1}{|a|} \int_{-\infty}^{+\infty} f_X\left(\frac{z-by}{a}\right) f_Y(y) dy$$

形象记忆掌握法: 这类公式特别有规律, 在形式上, 只要从 $z = ax + by$ 中解出 $y = \frac{z-ax}{b}$ 代换

$f_Y(y)$ 中的 y , 或者从 $z = ax + by$ 中解出 $x = \frac{z-by}{a}$ 代换 $f_X(x)$ 中的 x 即可。

如果使用雅可比变换公式求上述密度函数, 则较为简单, 则需人为添加恒等式 $X = X$, 则构成

$$\begin{cases} x = x \\ z = ax + by \end{cases} \Rightarrow \begin{cases} x = x \\ y = \frac{z-ax}{b} \end{cases} \Rightarrow J = \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -\frac{a}{b} & \frac{1}{b} \end{vmatrix} = \frac{1}{b}$$

从而 (X, Z) 的概率密度为 $f(x, z-x)|J| = \frac{1}{|b|} f\left(x, \frac{z-ax}{b}\right)$, $f_Z(z)$ 可以看成是 (X, Z) 的边缘分布, 故

$$f_Z(z) = \frac{1}{|b|} \int_{-\infty}^{+\infty} f_X(x) f_Y\left(\frac{z-ax}{b}\right) dx. \text{ 同理可以推得 } 3y \text{ 型。}$$

【备用模型直线平移法】

备选模型中, 已知两个随机变量的分布密度, 求它们线性组合 $Z = aX + bY$ 的分布函数密度 $f_Z(z)$, 可以直接使用备选模型公式。但是积分区间是分段的, 我们必须将 z 分割成不同的积分区间分别积分。关键和难点就是如何确定 z 的积分区间, 为此, 作者创立了平移法可以方便而清晰地解决这类题型。步骤如下

(1) 首先画出基准直线 $ax + by = 0$;

(2) 把基准直线逐一平移到正概率区域的全部边界点上, 从而得到正概率区域的分割边界线, 该直线与 x 轴的交点就是 x 方向的积分区间分段点, 与 y 轴的交点就是 y 方向的积分区间分段点。下面以例题来具体体现。

【两个一维分布构成的函数分布】

对于备用模型, 如果已知两个一元连续型随机变量的密度函数, 则先求出 $f(x, y)$, 再使用上述方法。

如果已知两个一元离散型随机变量的分布列, 或已知一个一元连续型随机变量的密度函数和一个一元离散型随机变量的分布列, 则使用全概率公式计算。

二、最值模型

最值模型分为并联模型和串联模型。要求 X, Y 独立, 并联模型事件表示为 $M = \text{Max}(X, Y)$, 串联模型事件表示为 $N = \text{Min}(X, Y)$ 。

2.1 离散最值模型题型具有定势

2.2 连续最值模型较为复杂

4 个必须掌握的事件关系:

$$\begin{aligned} \{\text{Max}(X, Y) > C\} &= (X > C) \cup (Y > C) \text{ 大大大并大} \\ \{\text{Max}(X, Y) \leq C\} &= (X \leq C, Y \leq C) = (X \leq C) \cap (Y \leq C) \text{ 小小小交小} \\ \{\text{Min}(X, Y) > C\} &= (X > C, Y > C) = (X > C) \cap (Y > C) \text{ 小大大交大} \\ \{\text{Min}(X, Y) \leq C\} &= (X \leq C) \cup (Y \leq C) \text{ 小小小并小} \end{aligned}$$

3 个常用的事件关系:

$$\begin{cases} (X+Y \leq 2C) \subset (X \leq C) \cup (Y \leq C) \\ M = \text{Max}(X, Y) = \frac{1}{2}(X+Y+|X-Y|) \\ N = \text{Min}(X, Y) = \frac{1}{2}(X+Y-|X-Y|) \end{cases}$$

n 个最值事件的概率计算公式

设随机变量 $X_1, X_2, \dots, X_n$ 相互独立, 它们的联合分布函数为 $F_{X_i}(x_i)$, 则

$$\begin{cases} M = \text{Max}\{X_1, X_2, \dots, X_n\} \Rightarrow F_{\text{Max}}(z) = P(M \leq z) = [F(z)]^n \Rightarrow f_{\text{Max}}(x) = nf(x)[F(z)]^{n-1} \\ N = \text{Min}\{X_1, X_2, \dots, X_n\} \Rightarrow F_{\text{Min}}(z) = P(N \leq z) = 1 - [1 - F(z)]^n \Rightarrow f_{\text{Min}}(x) = nf(x)[1 - F(z)]^{n-1} \end{cases}$$

三、商积模型

3.1 商模型

求商模型 $Z = \frac{aX}{bY}$ 的概率密度, 利用定义 $F_z(z) = P\{Z \leq z\} = P\left\{\frac{aX}{bY} \leq z\right\} = \iint_{\frac{ax}{by} \leq z} f(x, y) dx dy$ 麻烦, 现

利用雅可比变换公式求之。

$$\text{令 } u = y, v = \frac{ax}{by} \Leftrightarrow x = \frac{b}{a}uv, y = u \Rightarrow J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{b}{a}v & \frac{b}{a}u \\ 1 & 0 \end{vmatrix} = -\frac{b}{a}u$$

$$\Rightarrow f_z(z) = \int_{-\infty}^{+\infty} f\left(\frac{b}{a}zu, u\right) \left|\frac{b}{a}u\right| du = \left|\frac{b}{a}\right| \int_{-\infty}^{+\infty} f\left(\frac{b}{a}zy, y\right) |y| dy$$

如 X, Y 独立, 则

$$Z = \frac{aX}{bY} \Rightarrow f_z(z) = \left|\frac{b}{a}\right| \int_{-\infty}^{+\infty} f_x\left(\frac{b}{a}yz\right) f_y(y) |y| dy = \left|\frac{a}{bz^2}\right| \int_{-\infty}^{+\infty} f_x(x) f_y\left(\frac{ax}{bz}\right) |x| dx。$$

$$\text{类似有 } Z = \frac{bY}{aX} \Rightarrow f_z(z) = \left|\frac{b}{az^2}\right| \int_{-\infty}^{+\infty} f_x\left(\frac{by}{az}\right) f_y(y) |y| dy = \left|\frac{a}{b}\right| \int_{-\infty}^{+\infty} f_x(x) f_y\left(\frac{a}{b}xz\right) |x| dx$$

这是一个两个常用公式。希望读者能自己推导一次, 以助熟练应用。

【商模型旋转法】

商模型中, 已知两个独立的随机变量的分布密度求它们的商函数 $Z = \frac{aX}{bY}$ 的分布函数密度 $f_z(z)$, 可

以直接使用商模型公式 $Z = \frac{aX}{bY} \Rightarrow f_z(z) = \left|\frac{b}{az^2}\right| \int_{-\infty}^{+\infty} f_x\left(\frac{by}{az}\right) f_y(y) |y| dy$, 然而积分区间一般是分段的,

我们必须将 z 分割成不同的积分区间积分。问题关键和难点就是如何确定 z 的积分区间, 为此, 作者创立了旋转法可以方便而清晰地解决这类题型。旋转法步骤如下

(1) 首先画出基准直线 $\frac{ax}{by} = 1$;

(2) 把基准直线绕原点旋转到正概率区域的全部边界点上, 从而得到正概率区域的分割边界线, z 的取值范围由分割边界线确定, 该直线与 x 轴的交点就是 x 方向的积分区间分段点, 与 y 轴的交点就是 y 方向的积分区间分段点。

3.2 积模型

求积模型 $Z = XY$ 的概率密度。只要改写成 $Z = aXY = \frac{aX}{\frac{1}{Y}}$, 利用雅可比变换公式

$$\text{令 } u = \frac{1}{y}, \quad v = axy \Leftrightarrow x = \frac{1}{a}uv, \quad y = \frac{1}{u} \Rightarrow J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{v}{a} & \frac{u}{a} \\ -\frac{1}{u^2} & 0 \end{vmatrix} = \frac{1}{au}$$

$$\Rightarrow F_Z(z) = \frac{1}{|a|} \iint_{v \leq az} f\left(\frac{uv}{a}, \frac{1}{u}\right) |J| du dv = \frac{1}{|a|} \int_{-\infty}^z \left[\int_{-\infty}^{+\infty} f\left(\frac{uv}{a}, \frac{1}{u}\right) \frac{du}{|u|} \right] dv$$

$$\Rightarrow f_Z(z) = \frac{1}{|a|} \int_{-\infty}^{+\infty} f\left(\frac{zu}{a}, \frac{1}{u}\right) \frac{du}{|u|}$$

$$Z = aXY \Rightarrow f_Z(z) = \frac{1}{|a|} \int_{-\infty}^{+\infty} f\left(\frac{zu}{a}, \frac{1}{u}\right) \frac{du}{|u|} = \frac{1}{|a|} \int_{-\infty}^{+\infty} f\left(\frac{z}{ay}, y\right) \frac{dy}{|y|} = \frac{1}{|a|} \int_{-\infty}^{+\infty} f\left(x, \frac{z}{ax}\right) \frac{dx}{|x|}$$

【积模型曲线平移法】

在不为零的概率密度区域边界曲线的各个交点中, 能构成双曲线 $xy = x_0 y_0$ 的 $x_0 y_0$ 的可能值就是 z 的全部分解界点, 过每个分界点的双曲线之间“平行”, 即具有平移性。

第四章 随机变量的 6 大数字特征

6 大数字特征:

数学期望 EX ; 方差 DX ; 协方差 $Cov(X, Y)$ 或 σ_{XY} ; 相关系数 ρ_{XY} ; 矩; 协方差及其矩阵 Σ 。

一、数学期望

考研数学 4 种平均概念: 算数平均; 几何平均; 区间平均; 加权平均, 即概率平均, 也就是数学期望。

1.1 一维随机变量及函数的数学期望

$$(1) \text{ 离散型 } P\{X = x_k\} = p_k \Rightarrow EX = \sum_{k=1}^{\infty} x_k p_k \text{ or } \sum_{K=1}^{\infty} g(x_k) p_k$$

$$(2) \text{ 连续型 } EX = \int_{-\infty}^{+\infty} x f(x) dx \text{ or } \int_{-\infty}^{+\infty} g(x) f(x) dx, \quad f(x) \text{ 为 } X \text{ 的概率密度。}$$

1.2 二维随机变量函数的数学期望 $E[g(X, Y)]$

$$(1) \text{ 离散型 } P\{X = x_i, Y = y_j\} = p_{ij} \Rightarrow E[g(X, Y)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) p_{ij}$$

$$(2) \text{ 连续型 } E[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy$$

注意: $f(x, y)$ 是 (X, Y) 的密度函数, 而非 $g(X, Y)$ 的密度函数。

1.3 数学期望的常用结论

$$(1) E(C) = C; E(EX) = EX$$

$$(2) E(aX + bY) = aEX + bEY$$

$$(3) EXY = EX \cdot EY + E[(X - EX)(Y - EY)] \xrightarrow{X, Y \text{ 独立}} EXY = EX \cdot EY$$

$$(4) \varphi(x) \sim N(0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Rightarrow \int_{-\infty}^{+\infty} x \varphi(x) dx = \int_{-\infty}^{+\infty} (x - a) \varphi(x - a) dx = 0$$

二、方差

方差本质上是一个函数数学期望, 即是随机变量相对于中心值 EX 差的平方作为一个整体的数学期望。

$$2.1 \text{ 离散型 } DX = \sum_{k=1}^{\infty} (x_k - EX)^2 p_k$$

$$2.2 \text{ 连续型 } DX = \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx$$

通常使用公式 $DX = E(X - EX)^2 = EX^2 - E^2 X$ 计算方差, $\sqrt{DX} = \sigma_X$ 称为标准方差。

2.3 方差的常用结论

$$(1) D(C) = 0; D(EX) = 0$$

$$(2) D(\pm CX) = C^2 DX$$

$$(3) \quad D(X \pm Y) = DX + DY \pm 2E\{[X - EX][Y - EY]\} = DX + DY \pm 2Cov(X, Y) \\ = DX + DY \pm 2\sigma_{XY} = DX + DY \pm 2\rho_{XY}\sqrt{D(X)D(Y)}$$

$$(4) \quad X \text{ 与 } Y \text{ 独立} \quad D(aX \pm bY) = a^2 DX + b^2 DY$$

$$(5) \quad X \text{ 与 } Y \text{ 独立} \quad D(XY) = DXDY + DX(EY)^2 + DY(EX)^2$$

三、13 大分布的数学期望与方差

3.1 0-1 分布的数学期望和方差

0-1 分布又称为两点分布, 分布函数为 $P\{X = k\} = p^k(1-p)^{1-k} \sim B(1, p), k = 0, 1$

$$EX = p \quad DX = p(1-p)$$

3.2 二项分布的数学期望和方差

分布函数为 $P\{X = k\} = C_n^k p^k q^{n-k} \sim B(n, p)$, 为 n 个 0-1 分布之和。

$$EX = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = np \quad DX = D\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n D(X_i) = np(1-p)$$

3.3 泊松分布的数学期望和方差

分布函数为 $P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \sim P(\lambda) \quad (\lambda > 0)$, 当 $x = k = 0 \rightarrow P = e^{-\lambda}$

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

3.4 均匀分布的数学期望和方差

密度函数为 $f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其它} \end{cases} \sim U(a, b)$

$$EX = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}; \quad EX^2 = \frac{a^2 + ab + b^2}{3} \\ DX = EX^2 - (EX)^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

3.5 正态分布的数学期望和方差

密度函数为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \sim N(\mu, \sigma^2)$

$$EX = \mu \quad EX^2 = \mu^2 + \sigma^2 \quad DX = \sigma^2$$

$$E[\text{Max}(X_1, X_2)] = \mu + \frac{\sigma}{\sqrt{\pi}}; \quad E[\text{Min}(X_1, X_2)] = \mu - \frac{\sigma}{\sqrt{\pi}}, \\ E|X - Y| = E[\text{Max}(X_1, X_2) - \text{Min}(X_1, X_2)] = \frac{2\sigma}{\sqrt{\pi}}$$

3.6 指数分布的数学期望和方差

$$\text{密度函数为 } f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \quad (\lambda > 0) \\ 0, & x < 0 \end{cases} \sim E(\lambda)$$

$$EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} x\lambda e^{-\lambda x}dx = \frac{1}{\lambda}; \quad EX^2 = \int_{-\infty}^{+\infty} x^2\lambda e^{-\lambda x}dx = \frac{1}{\lambda^2} \int_{-\infty}^{+\infty} t^2 e^{-t}dt = \frac{2}{\lambda^2}$$

$$DX = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}$$

3.7 几何分布的数学期望和方差

$$\text{分布函数为 } P(X = k) = p(1-p)^{k-1} \sim G(p) \quad k=1, 2, \dots$$

$$EX = \frac{1}{p} \quad EX^2 = \frac{2-p}{p^2} \quad DX = E(X^2) - (E(X))^2 = \frac{1-p}{p^2}$$

3.8 超几何分布的数学期望和方差

$$\text{分布函数为 } P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \sim H(N, M, n)$$

$$EX = \frac{nM}{N} \quad DX = \frac{nM}{N} \left(1 - \frac{M}{N}\right) \left(\frac{N-n}{N-1}\right)$$

3.9 卡方分布的数学期望和方差

$$\text{分布密度函数为 } \chi^2(n). \quad \boxed{EX = n \quad DX = 2n}$$

3.10 T 分布的数学期望和方差

$$\text{分布密度函数为 } t(n). \quad \boxed{EX = 0 \quad DX = \frac{n}{n-2} \quad (n > 2)}$$

3.11 F 分布的数学期望和方差

$$\text{分布密度函数为 } F(n_1, n_2). \quad \boxed{EX = \frac{n_2}{n_2-2} \quad (n_2 > 2) \quad DX = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} \quad (n_2 > 4)}$$

3.12 二维均匀分布的数学期望和方差

$$f(x, y) \sim U(a, b; c, d) \Rightarrow \begin{cases} E(X, Y) = \left(\frac{a+b}{2}, \frac{c+d}{2}\right) \\ D(X, Y) = \left(\frac{(b-a)^2}{12}, \frac{(d-c)^2}{12}\right) \end{cases}$$

3.13 二维正态分布的数学期望和方差

$$f(x, y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho) \Rightarrow \begin{cases} E(X, Y) = (\mu_1, \mu_2) \\ D(X, Y) = (\sigma_1^2, \sigma_2^2) \end{cases}$$

四、二维随机变量的数字特征

4.1 数学期望

$$(1) \text{ 边缘分布离散型: } \boxed{EX = \sum_{i=1}^n x_i p_{i.}; \quad EY = \sum_{j=1}^n y_j p_{.j}}$$

(2) 边缘分布连续型: $EX = \int_{-\infty}^{+\infty} xf_X(x)dx; \quad EY = \int_{-\infty}^{+\infty} yf_Y(y)dy$

(3) 联合分布函数型:

$$E(G(X,Y)) = \sum_i \sum_j G(x_i, y_j) p_{ij}; \quad E(G(X,Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G(x,y) f(x,y) dx dy$$

4.2 方差

(1) 边缘分布离散型: $D(X) = \sum_{i=1}^n [x_i - E(X)]^2 p_i; \quad E(Y) = \sum_{i=1}^n [y_i - E(Y)]^2 p_{.j}$

(2) 边缘分布连续型: $D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f_X(x)dx; \quad E(Y) = \int_{-\infty}^{+\infty} [y - E(Y)]^2 f_Y(y)dy$

(3) 联合分布函数型:

$$D[G(X,Y)] = \sum_i \sum_j [G(x_i, y_j) - E(G(X,Y))]^2 p_{ij}$$

$$D[G(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [G(x,y) - E[G(X,Y)]]^2 f(x,y) dx dy;$$

(4) 随机变量的标准化方法 $Y = \frac{X - EX}{\sqrt{DX}} \sim N(0, 1)$

4.3 协方差与相关系数及矩

EX , EY 只反映了 X 和 Y 各自的平均值, DX , DY 反映的是 X 和 Y 各自偏离平均值的程度, 而协方差则反映 X 和 Y 之间的关系。

4.3.1 协方差

$$\text{Cov}(X, Y) = \sigma_{XY} = E[(X - EX)(Y - EY)]$$

4.3.2 协方差的性质

- 1) X 和 Y 独立或不相关, 则 $\text{Cov}(X, Y) = 0$;
- 2) $\text{Cov}(X, X) = E(X - EX)^2 = DX$
- 3) $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y) = ab\text{Cov}(Y, X)$
- 4) $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$

4.3.3 协方差的计算方法

1) 离散型 $\text{Cov}(X, Y) = \sum_i \sum_j (x_i - EX)(y_j - EY) p_{ij} \quad (p_{ij} = P\{X = x_i, Y = y_j\})$

2) 连续型 $\text{Cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - EX)(y - EY) f(x, y) dx dy$

3) 利用 $\text{Cov}(X, Y)$ 与 E 和 D 的关系是计算的主要方法

$$\text{Cov}(X, Y) = \sigma_{XY} = EXY - EXEY \quad \text{or} \quad D(X \pm Y) = DX + DY \pm 2\text{Cov}(X, Y)$$

$$4) \quad D\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n DX_i + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j)$$

4.3.4 相关系数

(1) 相关系数定义 $\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{DX} \cdot \sqrt{DY}} = \frac{\sigma_{XY}}{\sqrt{DXDY}}$ 描述 X, Y 的线性相关程度

相关系数本质上是一种线性逼近。考虑以 X 的线性函数 $a + bX$ 来近似表示 Y ，这种表示程度的好坏由下式 e 的最小值决定

$$e_{\text{Min}} = E[Y - (a + bX)]_{\text{Min}}^2 = (1 - \rho_{XY}^2)DY \quad \text{其中} \quad a = EY - bEX, \quad b = \frac{\text{Cov}(X, Y)}{DX}$$

(2) 相关系数的性质

- 相关系数 ρ_{XY} 反应了两个随机变量 X 和 Y 的线性关系， $|\rho_{XY}| = 1$ 的充要条件是使 $P\{Y = aX + b\} = 1$ ， $a \neq 0$ ，表示 X 和 Y 是完全的线性关系。
- $e_{\text{Min}} \geq 0 \Rightarrow |\rho_{XY}| \leq 1$ 。
- X 和 Y 独立，说明 X 和 Y 什么关系都没有，当然也不会有线性关系，从而 $\text{Cov}(X, Y) = 0 \Rightarrow \rho_{XY} = 0$ ； $\rho_{XY} = 0 \Rightarrow X$ 和 Y 不相关，但只能说明 X 和 Y 没有线性关系，但 X 和 Y 可能有非线性关系， X 和 Y 当然不一定独立。也就是说，独立必不相关，不相关不一定独立。
只有对正态分布和二值 0-1 分布而言，独立和不相关才是完全等价。
- 不相关的等价命题（均为充要条件）
 - 1) $\text{Cov}(X, Y) = 0$
 - 2) $\rho_{XY} = 0$
 - 3) $EXY = EX \cdot EY$
 - 4) $D(X \pm Y) = DX + DY$

4.3.5 矩和协方差矩阵

1) k 阶矩原点矩

$$E(x^k) = \sum_{i=1}^{\infty} x_i^k p_i = \int_{-\infty}^{+\infty} x^k f(x) dx$$

$$P\{X = x_i\} = p_i$$

2) k 阶中心矩

$$E[X - E(X)]^k = \sum_{i=1}^{\infty} [x_i - E(X)]^k p_i = \int_{-\infty}^{+\infty} [x - E(X)]^k f(x) dx$$

3) $k + \ell$ 阶混合矩

$$E\{[X - E(X)]^k [Y - E(Y)]^\ell\}$$

显然， EX 为 X 的一阶原点矩， DX 是 X 的二阶中心矩， $\text{Cov}(X, Y)$ 是 X, Y 的 1+1 阶混合中心矩，也就是说随机变量的全部数字特征最终都可以由矩来统一。

4.3.6 协方差矩阵

设 n 维随机变量 $(X_1, X_2, \dots, X_n)$ 的 $1+1$ 阶混合中心矩,

$$\sigma_{ij} = \text{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}$$

$$\text{则协方差矩阵定义为 } \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

由于 $\sigma_{ij} = \sigma_{ji}$, Σ 是一个对称矩阵, 它给出了 n 维随机变量的全部方差和协方差。

如对二维随机变量 (X_1, X_2) , 有四个二阶中心矩, 下面的 $\Sigma(X_1, X_2)$ 是重要考点。

$$\sigma_{ij} = \text{cov}(X_i, X_j)$$

$$\sigma_{11} = E\{[X_1 - E(X_1)]^2\} = \text{cov}(X_1, X_1) = DX_1$$

$$\sigma_{12} = E\{[X_1 - E(X_1)][X_2 - E(X_2)]\} = \text{cov}(X_1, X_2) = \sigma_{21}$$

$$\sigma_{22} = E\{[X_2 - E(X_2)]^2\} = \text{cov}(X_2, X_2) = DX_2$$

$$\Sigma(X_1, X_2) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} DX_1 & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & DX_2 \end{bmatrix} = \begin{bmatrix} DX_1 & \rho_{X_1 X_2} \sqrt{DX_1 DX_2} \\ \rho_{X_1 X_2} \sqrt{DX_1 DX_2} & DX_2 \end{bmatrix}$$

第五章 大数定理与中心极限定理

考点导读 3 大 2 中 1 不等 (3 个大数定理、2 个中心极限定理和一个不等式)。

一、切贝雪夫不等式

1. 切贝雪夫不等式及其应用范围

不管 X 属于何种分布, 只要 EX 和 DX 存在, 就可以估算出以 EX 为中心的对称区间上取值的概率。

即任给 $\varepsilon > 0$, 有 $P\{|X - EX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}$ 或 $P\{|X - EX| < \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2}$

证 明: 由积分比较定理可知:

$$\begin{aligned} DX &= \int_{-\infty}^{\infty} [x - EX]^2 f(x) dx \geq \int_{|x-EX| \geq \varepsilon} [x - EX]^2 f(x) dx \geq \int_{|x-EX| \geq \varepsilon} \varepsilon^2 f(x) dx \\ &= \varepsilon^2 \int_{|x-EX| \geq \varepsilon} f(x) dx = \varepsilon^2 P\{|X - EX| \geq \varepsilon\} \\ \Rightarrow P\{|X - EX| \geq \varepsilon\} &\leq \frac{DX}{\varepsilon^2} \Rightarrow 1 - P\{|X - EX| < \varepsilon\} \leq \frac{DX}{\varepsilon^2} \Rightarrow P\{|X - EX| < \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2} \end{aligned}$$

切贝雪夫不等式估计事件的概率是不精确的, 而且应注意两点:

(1) 要求随机变量 X 和 EX 之差构成的不等式能写成绝对值形式, 比如 $\{-20 < x - 50 < 21\}$ 事件就不能使用切贝雪夫不等式估计事件的概率; 当然还要知道 DX 。

对于任意事件 $a < X < b$, 仅当满足 $EX = \frac{a+b}{2}$ 时, 切贝雪夫不等式才可用, 并有

$$P\{a < X < b\} = P\left\{\frac{a-b}{2} < X - \frac{a+b}{2} < \frac{b-a}{2}\right\} = P\left\{\left|X - \frac{a+b}{2}\right| < \frac{b-a}{2}\right\} \geq 1 - \frac{4DX}{(b-a)^2}$$

(2) 要求是同一个随机变量, 即便对二维形式, 比如 $\{(X+Y) - E(X+Y)\}$ 也适用。

另外, 对于相关证明题, 其思想也是切贝雪夫不等式的证明模型, 主要是扩大积分区域与被积函数。

比如: 设 $f(x) > 0$, 且在 $[0, +\infty)$ 上单调非降, 则 $\forall \varepsilon > 0$, 有

$$P\{|X| \geq \varepsilon\} = \int_{|x| \geq \varepsilon} dF_X(x) \leq \int_{|x| \geq \varepsilon} \frac{f(|x|)}{f(\varepsilon)} dF_X(x) \leq \int_{-\infty}^{+\infty} \frac{f(|x|)}{f(\varepsilon)} dF_X(x) = \frac{1}{f(\varepsilon)} E[f(|x|)]$$

2. 依概率收敛的定义

设 a 是一个常数, X_n 为一随机变量序列, $\forall \varepsilon > 0$, $\exists P\{|X_n - a| < \varepsilon\} = 1$ 或 $P\{|X_n - a| \geq \varepsilon\} = 0$, 则称 $\{X_n\}$ 依概率收敛于 a , 记为 $X_n \xrightarrow{P} a$ 或 $\lim_{n \rightarrow +\infty} X_n = a(P)$ 。

如果 $z = g(x, y)$ 在 (a, b) 上连续, 且 $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, 则 $g(X_n, Y_n) \xrightarrow{P} g(a, b)$ 。

二、3 个大数定理

2.1 大数定理的定义

设随机变量序列 $\{X_n\}$ 的数学期望 $\{EX_n\}$ 存在 (不要求方差一定存在), 如果对任意 $\varepsilon > 0$, 有

$\lim_{n \rightarrow \infty} P\left(\left|\bar{X} - \frac{1}{n} \sum_{i=1}^n EX_i\right| < \varepsilon\right) = 1$, 则称 $\{X_n\}$ 服从大数定理。其中 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 。

2.2 大数定理的特征

体现一个“均”字。

大数定理中的随机变量、数学期望、方差（标准差）均是对“均”而言。如 $\bar{X}$ 和 $\frac{1}{n} \sum_{i=1}^n EX_i$

2.3 切比雪夫大数定理

设随机变量 $X_1, X_2, \dots, X_n$ 相互独立, 不要求同分布, 但数学期望和方差都存在 (注意 $DX = cn$, 当 $n \rightarrow +\infty$ 方差不存在), 则 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\left\{\left|\bar{X} - \frac{1}{n} \sum_{i=1}^n EX_i\right| < \varepsilon\right\} = 1。$$

2.4 辛钦大数定理

设随机变量 $X_1, X_2, \dots, X_n$ 相互独立同分布, 且具有相同的数学期望 $E(X_i) = \mu$ ($i = 1, 2, \dots, n$), 不求方差存在, 则 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\{|\bar{X} - \mu| < \varepsilon\} = 1$$

辛钦大数定理说明, 在大量 (一般要求 $n > 45$) 的测量值中, 算术平均当 n 无限增加时将接近期望值。

2.5 伯努利大数定理

设 Y_A 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 在每次试验中发生的概率, 则 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{Y_A}{n} - p\right| < \varepsilon\right\} = 1 \quad \text{或} \quad \lim_{n \rightarrow \infty} P\left\{\left|\frac{Y_A}{n} - p\right| \geq \varepsilon\right\} = 0$$

(1) $Y_A = X_1 + X_2 + \dots + X_n$, $X_1, X_2, \dots, X_n$ 相互独立, 且都服从概率为 p 的同一 (0-1) 分布; 当然 (0-1) 分布数学期望和方差也存在。

(2) 当 n 很大 (一般要求大于 45) 时, 事件发生的频率 $\frac{Y_A}{n}$ 具有稳定性, 且逼近于其概率, 这也是为什么在实际应用中, 常用频率来代替事件发生概率的原因。

(3) 显然, 它要求的条件最高: 不仅期望、方差存在, 而且同分布 (0-1)。

6. 三个大数定理的应用选择原则

大数定理提供了算术平均代替加权平均 (数学期望) 的理论根据, 适应于事件发生的平均值依概率收敛情形。如果能已知 EX , DX 都存在, 则使用切比雪夫大数定理; 如果仅知道 EX 存在, 而未知 DX 是否存在, 则使用辛钦大数定理; 如果是伯努利试验, 则使用伯努利大数定理。

三、2 个中心极限定理

3.1 中心极限的应用范围

(1) $X_1, X_2, \dots, X_n$ ($n > 45$) 独立同分布; (2) EX_n 和 DX_n 都存在, 且 $DX_n \neq 0$ ($n = 1, 2, \dots$)。

3.2 中心极限定理的特征

体现一个“和”字。

中心极限中的随机变量、数学期望、方差(标准差)均是对“和”而言。如 $\sum_{k=1}^n X_k$ 和 $n\mu$ 及 $\sqrt{n\sigma}$ 。

3.2.1 列维—林德伯格中心极限定理

又称独立同分布的中心极限定理。

设 $X_1, X_2, \dots, X_n, \dots$ 独立同分布, 且具有数学期望和大于零的方差, 设 $E(X_i) = \mu$, $D(X_i) = \sigma^2 \neq 0$,

则随机变量 $\sum_{i=1}^n X_k$ 的标准化量 $Y_n = \frac{\sum_{k=1}^n X_k - E\left(\sum_{k=1}^n X_k\right)}{\sqrt{D\left(\sum_{k=1}^n X_k\right)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma}}$ 的分布函数 $F_n(x)$ 满足

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P\left\{\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

$$(1) \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma}} \sim N(0,1), \quad \sum_{k=1}^n X_k \sim N(n\mu, n\sigma^2), \quad n \rightarrow \infty;$$

(2) 此处 Y_n 表达式中, 分子与分母可同乘以 $\frac{1}{n} \rightarrow Y_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, 正好对应标准化 $N(0, 1)$ 。

$$(3) P\left\{a \leq \sum_{k=1}^n X_k \leq b\right\} = P\left\{\frac{a - n\mu}{\sqrt{n\sigma}} \leq \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma}} \leq \frac{b - n\mu}{\sqrt{n\sigma}}\right\} \approx \Phi\left(\frac{b - n\mu}{\sqrt{n\sigma}}\right) - \Phi\left(\frac{a - n\mu}{\sqrt{n\sigma}}\right)$$

3.2.2 棣莫佛—拉普拉斯中心极限定理

设随机变量 $X_k \sim B(1, p)$, $\eta_n = \sum_{k=1}^n X_k \sim B(n, p)$ ($n=1, 2, \dots$) (二项分布也是要求 $X_1, \dots, X_n$ 相互独立, 同时隐含 $DX_k = \sigma^2 = np(1-p) \neq 0$), 则 $\forall x$, 随机变量 $\eta_n = \sum_{i=1}^n X_i$ 的标准化量 $\frac{\eta_n - np}{\sqrt{np(1-p)}}$

满足 $\lim_{n \rightarrow \infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$

$$\lim_{n \rightarrow \infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

$$(1) \frac{\eta_n - np}{\sqrt{np(1-p)}} \sim N(0,1), \quad \eta_n = \sum_{k=1}^n X_k \sim N[np, np(1-p)], \quad n \rightarrow \infty.$$

(2) 一般来说, $x \geq 3 \Rightarrow \Phi(x) \approx 1$; $x \leq -3 \Rightarrow \Phi(x) \approx 0$ 。

$$(3) P\{a \leq \eta_n \leq b\} = P\left\{\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right\} \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right).$$

2 个中心极限定理应用选择的方法

中心极限定理提供了任何备选事件发生的标准化量依概率收敛于 $N(0, 1)$ 的理论根据。

当 EX , DX 都存在, 且 $DX \neq 0$ 时, 如果是伯努利试验(离散型), 则使用莫佛—拉普拉斯中心极限定理;

一般型使用列维—林德伯格中心极限定理。

中心极限定理拓展性

中心极限定理的本质是当 n 充分大任何分布都服从正态分布 $N(\mu, \sigma^2)$, μ 为具体分布的数学期望, σ^2 为具体分布的方差。比如 $X_1, X_2, \dots, X_n$ 为来自均匀分布 $X \sim U(a, b)$ 的简单随机样本, 则当 n 充分大时:

$$\frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\frac{a+b}{2}, \frac{(b-a)^2}{12n}\right)。$$
 对于其余分布类推。

中心极限定理与切比雪夫不等式的关系

尽管切比雪夫不等式和中心极限定理都能给出估计值, 但是切比雪夫不等式只能粗略估计方差存在的随机变量在以数学期望为中心的对称区间上的概率, 而中心极限定理是用来近似计算, 因此其精度远比雪夫不等式所做的估计精确, 并且中心极限定理能够近似计算非对称区间上的概率, 因此, 只有在计算 $P\{|X - EX| \geq \varepsilon\}$ 比较困难时, 才使用比雪夫不等式。

第二篇 数理统计

第六章 数理统计的基本概念【数学 1, 3】

概率论中研究的随机变量分布都是假设已知的, 进而研究随机变量的性质、特点和规律, 而数理统计中研究的随机变量分布是未知或不完全知道的, 人们是通过对所研究的随机变量进行重复独立的观察, 得到许多观察值, 对这些数据进行分析, 从而推断随机变量的分布。

本章导读 3 大分布 8 类枢轴量。

一、总体和样本

如检测产品的某一项数量指标 (如研究 100 瓦灯泡的寿命)。需要检测的全体产品称为**总体** (如 60000 个 100 瓦的灯泡, 也可以是无限个), 记为 X , 总体中的每一元素称为**样品或个体** (如一个 100 瓦灯泡)。我们没有必要把全部 60000 个灯泡都测试, 所以, 需要从总体 (60000 个灯泡) 中随机抽取 n 个样品 (如取 50 个灯泡) 组成**样本**, 称为**抽样**, n 称为**样本容量**, 由于抽取的 50 各灯泡相对于 60000 个灯泡很小, 故放回与不放回抽样的区别可以忽略, 则样本能够看成是 n 个相互独立且分布相同的随机变量 (以后简称 “独立同”), 记为 $(X_1, X_2, \dots, X_{50})$, 即 50 维随机变量, 称为**简单随即样本**。显然, 测试前, $(X_1, X_2, \dots, X_{50})$ 就是一个 50 维随机变量, 测试完成后, $(X_1, X_2, \dots, X_{50})$ 就对应有一组具体值 $(x_1, x_2, \dots, x_{50})$, 称为**样本观察值**, 即**样本值**。

注意 $(X_1, X_2, \dots, X_{50})$ 是一次所选择的需要测试的灯泡, 是随机抽取的, 对应于每一次的 $(X_1, X_2, \dots, X_{50})$, 而测试样本值 $(x_1, x_2, \dots, x_{50})$ 是确定的, 每次抽取了 50 个全部测试一次, 至于究竟需要抽取多少次, 则由测试要求决定。

如抽取样本 $(X_1, X_2, \dots, X_n)$ 10 次, 每次就相当于一个 50 维随机向量, 10 个随机向量称**样本空间**, 记为 Ω , 一次测试所得的一组样本观察值 $(x_1, x_2, \dots, x_n)$ 是 Ω 中的一个**样本点**, 容量为 n 的简单随机样本的数字特征及分布就代表了总体的特性, 例如, 研究 50 个灯泡的寿命就能代表 60000 个灯泡这个总体 X 的寿命。

注意若 $X_1, X_2, \dots, X_n$ 相互独立, 则 $Y_1 = f(X_1, X_2, \dots, X_n)$ 和 $Y_2 = g(X_1, X_2, \dots, X_n)$ 也相互独立。

二、样本函数和样本统计量

2.1 统计量

不含任何未知参数的 $g(X_1, X_2, \dots, X_n)$ 函数形式称为样本统计量, $g(x_1, x_2, \dots, x_n)$ 为相应样本值; 含任何未知参数的 $g(X_1, X_2, \dots, X_n)$ 就称为样本函数。统计量与样本函数一般在测试前后可以相互转化角色。

如把样本观测值 $(x_1, x_2, \dots, x_n)$ 按照顺序排列成 $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$, 其中 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, 记取值为 $x_{(k)}$ 的样本分量为 $X_{(k)}$, 则称 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 为顺序统计量。

如最大顺序统计量与最小顺序统计量为:

$$X_{(1)} = \text{Max}\{X_1, X_2, \dots, X_n\} \Rightarrow F_U(u) = [F_X(u)]^n \Rightarrow f_U(u) = n f_X(u) [F_X(u)]^{n-1}$$

$$X_{(n)} = \text{Min}\{X_1, X_2, \dots, X_n\} \Rightarrow F_V(v) = 1 - [1 - F_X(v)]^n \Rightarrow f_V(v) = n f_X(v) [1 - F_X(v)]^{n-1}$$

2.2 样本矩

也是一种样本函数, 注意 n 是变量, $\bar{X}$ 是随机变量。注意样本矩和数字特征中的矩定义不同。

$$(1) \text{ 原点矩: } A_k = \frac{1}{n} \sum_{i=1}^n X_i^k; \quad (2) \text{ 中心矩: } B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

2.3 常用样本函数

2.3.1 样本均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \text{ 为样本一阶原点矩。}$$

2.3.2 样本方差与标准差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right]; \quad S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right]}$$

注意区别于数字特征中的方差 σ^2 , σ^2 只是某一个随机变量 X_i 的方差, 而 S^2 是 n 个 X_i 的联合分布函数的方差。另外, 严格地说, S^2 不是矩。

2.3.3 二阶样本中心矩

$$B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right] = \frac{n-1}{n} S^2 \text{ 与样本方差 } S^2 \text{ 是不同的概念。}$$

相应统计量的观察值形式同上。

2.4 经验 (样本) 分布函数

设样本 $(X_1, X_2, \dots, X_n)$ 是取自总体 X , 则经验分布函数定义为

$$F_n(x) = \frac{1}{n} \{ (X_1, X_2, \dots, X_n) \text{ 中小于或等于 } x \text{ 的个数 } k \} = \frac{k}{n}$$

三、三种抽样分布

3.1 卡方分布

设 $\{X_i\}$ 独立同, $X_i \sim N(0,1)$, 则 $Z = \sum_{i=1}^n X_i^2 \sim \chi^2(n)$ 。

$$\text{其密度函数为 } f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}, \text{ 其中 } \Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

$$\Rightarrow \Gamma(\alpha+1) = \alpha \Gamma(\alpha), \Gamma(n+1) = n!, \Gamma(1) = 1, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$(1) \text{ 可加性 } \chi^2(n_1) + \chi^2(n_2) + \dots \sim \chi^2(n_1 + n_2 + \dots)$$

(2) 上分位点: $\chi_{\alpha}^2(n)$ 定义为 $\chi^2(n)$ 分布的分位点, 则 $P\{\chi^2(n) \geq \chi_{\alpha}^2(n)\} = \int_{\chi_{\alpha}^2(n)}^{+\infty} f_{\chi^2(n)}(x) dx = \alpha$ 。

上分位点的特点是对应分布函数图形 $\chi_{\alpha}^2(n)$ 点的右边面积, 对 $t(n)$ 和 $F(n_1, n_2)$ 分布有类似的结论。

(3) 卡分布的期望与方差 $E\chi^2(n) = n; D\chi^2(n) = 2n$

证明: 由于 $X_i \sim N(0, 1) \Rightarrow E(X_i) = 0; DX_i = 1$

$$EX_i^2 = D(X_i) + (EX_i)^2 = 1 \quad (i=1, 2, \dots, n); EX_i^4 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = 3$$

$$DX_i^2 = EX_i^4 - (EX_i^2)^2 = 3 - 1 = 2; E\chi^2(n) = E\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n EX_i^2 = n; D\chi^2(n) = D\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n DX_i^2 = 2n$$

$$(4) f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases} \Rightarrow \chi^2(2) = \begin{cases} \frac{1}{2} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases} \sim E\left(\frac{1}{2}\right)。$$

(5) 常用样本函数的数学期望和方差

卡方分布的一个常年考点是利用 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ (暂不要求读者证明) 间接计算样本函数的数学期望和方差, 其根据是 $E\frac{(n-1)S^2}{\sigma^2} = E\chi^2(n-1) = n-1; D\frac{(n-1)S^2}{\sigma^2} = D\chi^2(n-1) = 2(n-1)$ 。

$$X \sim N(\mu, \sigma^2) \Rightarrow EX = \mu; EX^2 = \mu^2 + \sigma^2; DX = \sigma^2$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \Rightarrow E\bar{X} = \mu; E\bar{X}^2 = \mu^2 + \frac{\sigma^2}{n}; D\bar{X} = \frac{\sigma^2}{n}$$

$$ES^2 = E\left[\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)\right] = \frac{1}{n-1} \left(\sum_{i=1}^n EX_i^2 - nE\bar{X}^2\right) = \frac{1}{n-1} \left[n(\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right] = \sigma^2$$

$$DS^2 = \frac{\sigma^4}{(n-1)^2} D\frac{(n-1)S^2}{\sigma^2} = \frac{\sigma^4}{(n-1)^2} D\chi^2(n-1) = \frac{2\sigma^4}{n-1}$$

$$E(B_2) = E(S^{*2}) = E\left(\frac{n-1}{n} S^2\right) = \frac{n-1}{n} \sigma^2; D(B_2) = D\left(\frac{n-1}{n} S^2\right) = \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n}$$

$$DX^2 = \sigma^4 D\left(\frac{X-\mu}{\sigma} + \frac{\mu}{\sigma}\right)^2 = \sigma^4 D\left(\left(\frac{X-\mu}{\sigma}\right)^2 + 2\left(\frac{\mu}{\sigma}\right)\left(\frac{X-\mu}{\sigma}\right) + \left(\frac{\mu}{\sigma}\right)^2\right)$$

$$= \sigma^4 D\left(\chi^2(1) + 2\left(\frac{\mu}{\sigma}\right)N(0, 1)\right) = \sigma^4 \left(2 + \frac{4\mu^2}{\sigma^2}\right) = 2\sigma^4 + 4\mu^2\sigma^2$$

$$\begin{aligned}
 D\bar{X}^2 &= \frac{\sigma^4}{n^2} D\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} + \frac{\mu}{\sigma/\sqrt{n}}\right)^2 = \frac{\sigma^4}{n^2} D\left[\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 + 2\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)\frac{\mu}{\sigma/\sqrt{n}} + \left(\frac{\mu}{\sigma/\sqrt{n}}\right)^2\right] \\
 &= \frac{\sigma^4}{n^2} D\chi^2(1) + \frac{\sigma^4}{n^2} \times \frac{4n\mu^2}{\sigma^2} DN(0, 1) = \frac{2\sigma^4}{n^2} + \frac{4n\mu^2\sigma^2}{n^2} = \frac{2\sigma^4 + 4n\mu^2\sigma^2}{n^2} \\
 \mu = 0 &\Rightarrow D\bar{X}^2 = \frac{2\sigma^4}{n^2}
 \end{aligned}$$

$$\begin{aligned}
 S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \Rightarrow \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{(n-1)S^2}{n} + \bar{X}^2 \\
 D\bar{X}^2 &= D\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) = D\left(\frac{(n-1)S^2}{n}\right) + D\bar{X}^2 = \frac{\sigma^4}{n^2} D\left(\frac{(n-1)S^2}{\sigma^2}\right) + \frac{\sigma^4}{n^2} D\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} + \frac{\mu}{\sigma/\sqrt{n}}\right)^2 \\
 &= \frac{\sigma^4}{n^2} D\chi^2(n-1) + \frac{\sigma^4}{n^2} D\left[\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 + 2\left(\frac{\mu}{\sigma/\sqrt{n}}\right)\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right) + \left(\frac{\mu}{\sigma/\sqrt{n}}\right)^2\right] \\
 &= \frac{\sigma^4}{n^2} D\chi^2(n-1) + \frac{\sigma^4}{n^2} D\chi^2(1) + \frac{\sigma^4}{n^2} \times \frac{4n\mu^2}{\sigma^2} = \frac{\sigma^4}{n^2} \times 2(n-1) + \frac{2\sigma^4}{n^2} + \frac{4\mu^2\sigma^2}{n} = \frac{2\sigma^4 + 4\mu^2\sigma^2}{n} \\
 \mu = 0 &\Rightarrow D\bar{X}^2 = \frac{2\sigma^4}{n}
 \end{aligned}$$

$$S_w^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} \Rightarrow ES_w^2 = \sigma^2$$

对于其它分布, 只要将相应的“等价”数字代入 $E\bar{X} = \mu$; $D\bar{X} = \frac{\sigma^2}{n}$; $ES^2 = \sigma^2$; $DS^2 = \frac{2\sigma^4}{n-1}$

类比求出 $E(\bar{X})$, $D(\bar{X})$, $E(S^2)$, $D(S^2)$, 比如

- $X \sim P(\lambda) \Rightarrow E\bar{X} = \lambda$; $D\bar{X} = \frac{\lambda}{n}$, $ES^2 = \lambda$, $DS^2 = \frac{2\lambda^2}{n-1}$
- $X \sim E(\lambda) \Rightarrow E\bar{X} = \frac{1}{\lambda}$; $D\bar{X} = \frac{1}{n\lambda^2}$, $ES^2 = \frac{1}{\lambda^2}$, $DS^2 = \frac{2}{(n-1)\lambda^2}$
- $X \sim B(n, p) \Rightarrow E\bar{X} = np$; $D\bar{X} = p(1-p)$, $ES^2 = np(1-p)$, $DS^2 = \frac{2[np(1-p)]^2}{n-1}$
- $X \sim U(a, b) \Rightarrow E\bar{X} = \frac{a+b}{2}$; $D\bar{X} = \frac{(b-a)^2}{12n}$, $ES^2 = \frac{(b-a)^2}{12}$, $DS^2 = \frac{(b-a)^4}{72(n-1)}$

量纲法求复合统计量的分布函数。

3 种抽样源正态, 量纲法则判类型; 根据定义凑模式, 标准变量容量值。

3.2 T 分布

设 $\{X_i\}$ 独立同分布, $X_i \sim N(0,1)$, $Y \sim \chi^2(n)$; X 和 Y 独立, 则 $T = \frac{X}{\sqrt{Y/n}} \sim t(n)$

$$\text{密度函数为 } f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < +\infty$$

$$(1) \quad t \text{ 分布密度函数当 } n \rightarrow \infty \Rightarrow f_{t(n)}(x) \rightarrow N(0,1)$$

$$(2) \quad \text{上分位点: } t_\alpha(n) \text{ 定义为 } t(n) \text{ 分布的分位点, 则 } P\{t(n) \geq t_\alpha(n)\} = \int_{t_\alpha(n)}^{+\infty} f_t(x) dx = \alpha$$

$$(3) \quad \text{性质: } T \text{ 分布具有对称性, } t_{1-\alpha}(n) = -t_\alpha(n); \quad n > 45 \text{ 时, } t_\alpha(n) \approx Z_\alpha, \quad t_{0.5}(n) = 0$$

$$\text{其中 } Z_\alpha \text{ 为标准正态分布的下分位点, 即 } P\{\Phi(x) \leq Z_\alpha\} = \int_{-\infty}^{Z_\alpha} N(0,1) dx = \alpha$$

$$(4) \quad ET = 0, \quad DT = \frac{n}{n-2}, \quad n > 2$$

3.3 F 分布

$$\text{设 } X, Y \text{ 相互独立, } X \sim \chi^2(m); \quad Y \sim \chi^2(n); \quad \text{则 } F = \frac{X/m}{Y/n} \sim F(m, n)$$

$$\text{密度函数为 } f(x) = \begin{cases} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$\text{特别地, } m=n=1 \Rightarrow F(1,1) = \frac{1}{\pi(1+x)\sqrt{x}}.$$

$$F_\alpha(n,m) \text{ 定义为 } F(n,m) \text{ 分布的分位点, 则 } P\{F \geq F_\alpha(n,m)\} = \int_{F_\alpha(n,m)}^{+\infty} f_F(x) dx = \alpha$$

性 质

$$(1) \quad F_{1-\alpha}(n,m) = \frac{1}{F_\alpha(m,n)}; \quad F(n,m) = \frac{1}{F(m,n)}$$

$$(2) \quad X \sim t(n) \Rightarrow X^2 \sim F(1,n); \quad X \sim F(n,m) \Rightarrow \frac{1}{X} \sim F(m,n)$$

$$(3) \quad EF(n_1, n_2) = \frac{n_2}{n_2-2}, n_2 > 2; \quad DF(n_1, n_2) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}, n_2 > 4$$

$$(4) \quad n_1 \geq 3 \Rightarrow F \text{ 分布密度函数 } f(x) \text{ 的极大值点 } x = \frac{(n_1-2)n_2}{n_1(n_2+2)}.$$

四、数理统计中 8 大标准枢轴量分布结论与推导

枢轴量是样本函数和统计量的总称。

4.1 单个正态总体

设 $\{X_n\} \sim N(\mu, \sigma^2)$ 为一系列简单随机样本, 则有

(1) 若 σ 已知, 需要估计 μ 的范围, 则使用枢轴量 $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$

(2) 若 μ 未知, 需要估计 σ 的范围, 则使用枢轴量

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2(n-1); \text{ 且 } \bar{X} \text{ 与 } S^2 \text{ 独立} \quad (\bar{X} \text{ 是随机变量})$$

(3) 若 σ 未知, 需要估计 μ 的范围, 则使用枢轴量 $\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$

(4) 若 μ 已知, 需要估计 σ 的范围, 则使用枢轴量 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$

4.2 两个正态总体

X 和 Y 独立同分布。

$$X \sim N(\mu, \sigma_1^2), \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i; \quad Y \sim N(\mu_2, \sigma_2^2), \quad \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i$$

$$S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2$$

(5) 若 σ_1^2, σ_2^2 已知, 需要估计 $\mu_1 - \mu_2$ 的范围, 则使用枢轴量 $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$

(6) 若 σ_1^2, σ_2^2 未知, 但 $\sigma_1 = \sigma_2 = \sigma$ 时, 需要估计 $\mu_1 - \mu_2$ 的范围, 则使用枢轴量

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2) \xrightarrow{\text{当 } m=n} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2 + S_2^2}{n}}} \sim t(2n-2)$$

(7) 如 μ_1, μ_2 已知, 需要估计 $\frac{\sigma_1^2}{\sigma_2^2}$ 的范围, 则使用枢轴量 $\frac{\frac{1}{n} \sum_{i=1}^n (X_i - \mu_1)^2}{\frac{1}{m} \sum_{i=1}^m (Y_i - \mu_2)^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F(n, m)$

(8) 如 μ_1, μ_2 未知, 需要估计 $\frac{\sigma_1^2}{\sigma_2^2}$ 的范围, 则使用枢轴量 $\frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F(n-1, m-1)$

8 大枢轴量贯穿考研数学数理统计的全部考点, 记忆口诀为:

单正态 4 分布。知方差 标准型;
未知方差 t 差 1; 知期望 卡平方 (χ^2) ;
未知期望卡 (χ^2) 减 1。双正态 估差比;
知方差 与单同; 未知方差 t 减 2;
知期望 用 F ; 未知期望 F 差 1。
含 S , 具特征; 每个容量减去 1。

第七章 参数估计 [13]

本章导读 统计推断有两类基本问题, 核心是八大枢轴量。

(1) 当分布函数形式为已知, 如为 $F(x, \theta_1, \theta_2, \dots, \theta_k)$, 但分布函数中的部分参数为未知时 (如 x 已知, $\theta_1, \theta_2, \dots, \theta_k$ 未知), 我们需要根据样本所提供的信息对总体进行统计推断, 估计其未知的参数, 如数学期望、方差等, 这类问题称为参数估计, 分为点估计与区间估计两种方法。根据样本是以随机变量 X 还是观察值 x 形式给出, 参数估计的结果又有估计量和估计值之分, 对应的字母有大小写之分。参数估计中, 用来自总体的一个样本或样本观测值去近似代替总体分布中的未知参数值, 称为点估计的估计量或估计值; 点估计量或估计值接近真值有多大的把握? 很难说, 比如, 估计一个人的身高, 如果你确切说他一米七, 很可能不符合事实, 如果你说他一米七左右, 可能更精确, 即估计时给出一定的误差范围, 而真值极大可能在这一误差区域内, 这种估计方法称为区间估计, 它是根据 8 个枢轴量之一来判断。

(2) 当分布函数形式为未知, 或只知道分布函数的形式, 参数为未知时, 我们也可以先假设它的形式或参数值, 然后再根据 8 个枢轴量之一, 根据小概率原理和样本值检验其假设的真实, 这一过程称为假设检验, 这是第八章的内容。

一、点估计 [13]

点估计分为矩估计与极大似然法两种。

1.1 矩估计法

矩估计的原则是【同型估计】, 即总体的矩是几阶, 就使用几阶的样本矩形式来估计。考纲范围内只要求掌握一个 (单参数, 如数学期望) 或两个 (双参数, 如数学期望和方差) 未知参数的情形。矩估计具有幂增传递性, 如 $\bar{X}$ 是 μ 的矩估计, 则 $\bar{X}^2$ 是 μ^2 的矩估计, 反之不正确。用矩法构造未知参数估计量的 3 类定势如下:

由于随机变量的数字特征都可以使用矩形式表示, 故一般先计算数学期望和方差, 再用同阶矩形式联立。

(1) 单参数矩估计法

对于单参数, 先计算总体 EX (假设数学期望为 $\mu \neq 0$), 再用样本一原点矩 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 替换 μ , 即 $\bar{X} = EX = \hat{\mu}$; 当 $EX = 0$ 时, 继续求出总体的 EX^2 , 再用样本二原点矩 $\bar{X}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ 替换, 求出参数值即得结论。

(2) 双参数矩估计法

对于两个参数, 先计算样本总体 EX 和 EX^2 , 再用样本一二阶原点矩

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ 和 } \overline{X^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 \text{ 分别替换总体 } EX \text{ 和 } EX^2; \text{ 列出方程组 } \begin{cases} EX = \frac{1}{n} \sum_{i=1}^n X_i \\ EX^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}, \text{ 求解参数的估计值。}$$

矩估计量的不变性原理 若 $\hat{\theta}$ 是 θ 的矩估计量, 且 $g(x)$ 为连续函数, 则 $g(\hat{\theta})$ 为 $g(\theta)$ 的矩估计量。

矩估计适合未知总体的分布的情形, 缺陷是估计值不唯一; 如果知道分布的参数形式, 则需要使用最大似然法。

(3) 方差的估计法

由于有时 DX 比 EX^2 更容易求, 所以在对两个参数进行矩估计时, 可以先用含待估计的两个参数的式子表示 EX 和 DX , 然后反解出参数, 即将参数用 EX 和 DX 表示, 再将 EX 用样本的一阶原点矩 $\frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ 代换, 将 DX 用样本的二阶中心矩 $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 代换即可。原理如下:

$$\begin{aligned} DX &= EX^2 - (EX)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 = \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) = \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - 2n\bar{X} + n\bar{X}^2 \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - 2n \cdot \bar{X} \cdot \frac{1}{n} \sum_{i=1}^n X_i + n\bar{X}^2 \right) = \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - 2n \cdot \frac{1}{n} \sum_{i=1}^n X_i \bar{X} + n\bar{X}^2 \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n 2X_i \bar{X} + \sum_{i=1}^n \bar{X}^2 \right) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

故随机变量方差 DX (注意不是样本方差 S^2) 可以是使用中心矩 $DX = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 来估计的。

如果题中给出了样本值, 则称为矩估计值; 如果题中未给出样本值, 则称为矩估计量。

【例】参见浙江大学三版 P178 例 2, 例题 3。

【例】求几何分布 $G(p)$ 和泊松分布 $P(\lambda)$ 的估计量。

【解】 $G(p) = P\{X=k\} = p(1-p)^{k-1} (k=1, 2, \dots)$

$$EX = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \left(-\sum_{k=1}^{\infty} (1-p)^k \right)' = \frac{1}{p} = \bar{X} \Rightarrow \hat{p} = \frac{1}{\bar{X}}.$$

$$P(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} (k=1, 2, \dots)$$

$$EX = \lambda = \bar{X} \Rightarrow \hat{\lambda} = \bar{X} \text{ or } DX = \lambda = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ or } \hat{\lambda} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2$$

可见, 矩估计是不唯一的。

【例】设总体 X 的密度函数 $f(x, \theta) = \begin{cases} \frac{2(\theta-x)}{\theta^2}, & 0 < x < \theta \\ 0, & \text{other} \end{cases}$, 求 θ 的矩估计。

【解】 $EX = \int_0^{\theta} \frac{2(\theta-x)}{\theta^2} x dx = \frac{\theta}{3} = \bar{X} \Rightarrow \hat{\theta} = 3\bar{X}.$

【例】设总体 X 分布律为

X	0	1	2	3
p	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中 $\theta \in (0, 1/2)$ 未知, 已知总体 X 有样本值: 3, 1, 3, 0, 3, 1, 2, 3, 求 θ 的矩估计值。

$$\text{【解】 } EX = 0 \times \theta^2 + 1 \times 2\theta(1-\theta) + 2 \times \theta^2 + 3 \times (1-2\theta) = 3 - 4\theta = \bar{x} = \frac{3+1+3+0+3+1+2+3}{8} = 2$$

$$\Rightarrow \hat{\theta} = \frac{1}{4} = 0.25。$$

1.2 最大似然法

最大似然法的基本思想是: “当用联合分布函数来估计总体参数时, 应使得当参数取一定值时, 使所观察到的样本出现的概率最大”。具体计算定势如下:

(1) 写出似然函数 $L = L(x; \theta_1, \theta_2)$ 。连续型似然函数形式为 $L = \prod_{i=1}^n f(x_i; \theta_1, \theta_2)$ 为联合分布密度函数;

离散型然函数形式为 $L = \prod_{i=1}^n P(x_i; \theta_1, \theta_2)$ 为联合分布律

(2) 为方便计算, 取对数 $\ln L = \sum_{i=1}^n \ln f$ 或 $\ln L = \sum_{i=1}^n \ln P$

(3) 当偏导存在时, 求解方程组 $\frac{\partial \ln L}{\partial \theta_i} = 0 \quad (i=1, 2, \dots, k)$ 解出 θ_i , 然后用 $\hat{\theta}_i$ 取代 θ_i 。如果偏导不

存在, 则使用极大似然法的基本思想讨论结果, 也就是使 $L = L(x; \theta_1, \theta_2)$ 在 $(\hat{\theta}_1, \hat{\theta}_2)$ 处取得最大值。

最大似然估计量的不变性原理

设 $\hat{\theta}$ 是 θ 的最大似然估计量, $g(\theta)$ 有单值反函数, 则 $g(\hat{\theta})$ 是 $g(\theta)$ 极大似然估计量。

$$\text{比如 } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}。$$

【例】设总体 X 分布律为

X	0	1	2	3
p	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中 $\theta \in (0, 1/2)$ 未知, 已知总体 X 有样本值: 3, 1, 3, 0, 3, 1, 2, 3, 求 θ 的最大似然估计值。

【解】对于给定的 8 个样本值, 有 1 个 0, 2 个 1, 2 个 1, 4 个 3, 故

$$L(\theta) = \theta^2 \cdot [2\theta(1-\theta)]^2 \cdot \theta^2 \cdot (1-2\theta)^4 = 4\theta^6 (1-\theta)^2 (1-2\theta)^4$$

$$\ln L = \ln 4 + 6 \ln \theta + 2 \ln(1-\theta) + 4 \ln(1-2\theta) \Rightarrow \frac{\partial \ln L}{\partial \theta} = \frac{6}{\theta} - \frac{2}{1-\theta} - \frac{8}{1-2\theta} = 0 \Rightarrow \theta = \frac{7 \pm \sqrt{13}}{12}$$

$$\theta = \frac{7 + \sqrt{13}}{12} > \frac{1}{2} \text{ 舍去, 故 } \theta \text{ 的最大似然估计值 } \hat{\theta} = \frac{7 - \sqrt{13}}{12} \approx 0.28。$$

对比矩估计, 二者是很接近的, 但不能确定谁更精确。

【例】求总体 $X \sim B(n, p)$, p 为未知常数, 求其最大似然估计量。

【解】设 $x_1, x_2, \dots, x_n$ 为 $X_1, X_2, \dots, X_n$ 的一个样本值, 总体 $X \sim B(n, p)$ 的概率密度函数为

$$f(x, p) = P\{X = x\} = C_n^x p^x (1-p)^{n-x} \quad (x=0, 1, \dots, n)$$

$$L(p) = L(x_1, x_2, \dots, x_n; p) = \prod_{i=1}^n C_n^{x_i} p^{x_i} (1-p)^{n-x_i} = \left(\prod_{i=1}^n C_n^{x_i} \right) \left(p^{\sum_{i=1}^n x_i} \right) \left((1-p)^{\sum_{i=1}^n (n-x_i)} \right)$$

$$\ln L = \sum_{i=1}^n \ln C_n^{x_i} + \ln p \cdot \sum_{i=1}^n x_i + \ln(1-p) \cdot \sum_{i=1}^n (n-x_i)$$

$$\Rightarrow \frac{\partial \ln L}{\partial p} = \frac{1}{p} \sum_{i=1}^n x_i + \frac{1}{1-p} \sum_{i=1}^n (n-x_i) = 0 \Rightarrow p = \frac{1}{n} \sum_{i=1}^n x_i = \frac{\bar{X}}{n} \Rightarrow \hat{p} = \frac{\bar{X}}{n}$$

【例】求总体 $X \sim N(\mu, \sigma^2)$ 的矩估计量和最大似然估计量。

【解】(1) 矩估计量计算

$$\begin{cases} EX = \mu = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \\ EX^2 = \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases} \Rightarrow \begin{cases} \mu = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \\ \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{cases}$$

当然, 也可以根据本章前面的理论直接给出方差的估计量。

假设题中有三个未知常数, 则再使用 $EX^3 = \frac{1}{n} \sum_{i=1}^n X_i^3$, 依次类推。

(2) 最大似然估计量计算

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

$$L(\mu, \sigma^2) = L(x_1, x_2, \dots, x_n; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(\sqrt{2\pi\sigma})^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2}$$

$$\ln L = -\frac{n}{2} (\ln(2\pi) + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\Rightarrow \begin{cases} \frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{cases}$$

【例 14】设总体 X 的概率密度函数为 $f(x, a, b) = \begin{cases} a, & -1 < x < 0 \\ b, & 0 \leq x < 1 \\ 0, & \text{other} \end{cases}$, X 取得样本值:

$-0.8, -0.6, -0.5, -0.2, -0.1, 0.4, 0.3, 0.5$, 求 a 的矩估计值与极大似然法估计值。

$$\text{【解】 } a+b=1 \Rightarrow b=1-a \Rightarrow f(x, a, b) = \begin{cases} a, & -1 < x < 0 \\ 1-a, & 0 \leq x < 1 \\ 0, & \text{other} \end{cases}$$

$$EX = \int_{-1}^0 axdx + \int_0^1 (1-a)xdx = \frac{1}{2} - a$$

$$\bar{x} = \frac{1}{8}(-0.8 - 0.6 - 0.5 - 0.2 - 0.1 + 0.4 + 0.3 + 0.5) = -\frac{1}{8} \Rightarrow EX = \bar{x} \Rightarrow \frac{1}{2} - \hat{a} = -\frac{1}{8} \Rightarrow \hat{a} = \frac{5}{8}$$

$$L(a) = a^5(1-a)^3 \Rightarrow \ln L(a) = 5\ln a + 3\ln(1-a) \Rightarrow \frac{\partial L(a)}{\partial a} = \frac{5}{a} - \frac{3}{1-a} = 0 \Rightarrow \hat{a} = \frac{5}{8}$$

如类似均匀分布的联合密度函数不可导或为常数, 则根据最大似然估计的原则, 直接取满足似然函数 L 最大值的参数值, 相关例题请研究浙大三版 P183 例题 6。

常用标准分布的点估计结论:

分布	矩估计	最大似然估计
$B(1, p)$	$\hat{p} = \bar{X}$	$\hat{p} = \bar{X}$
$B(n, p)$	$\hat{p} = \bar{X} / n$	$\hat{p} = \bar{X} / n$
$P(\lambda)$	$\hat{\lambda} = \bar{X}$	$\hat{\lambda} = \bar{X}$
$E(\lambda)$	$\hat{\lambda} = 1 / \bar{X}$	$\hat{\lambda} = 1 / \bar{X}$
$G(p)$	$\hat{p} = 1 / \bar{X}$	$\hat{p} = 1 / \bar{X}$
$U(a, b)$	$\hat{a} = \bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$ $\hat{b} = \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$	$\hat{a} = \min_i \{X_i\}$ $\hat{b} = \max_i \{X_i\}$
$N(\mu, \sigma^2)$	$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$	$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

1.3 点估计的 3 大评价标准 [1]

1.3.1 无偏性 (即没有系统误差)

如 $E(\hat{\theta}) = \theta$, 则 $\hat{\theta}$ 为 θ 的无偏估计量, 其中 θ 为原分布的真值。对于正态总体 $N(\mu, \sigma^2)$, 由于 $E(\bar{X}) = \mu$, 故是无偏性。而 $E(B_2) = \frac{n-1}{n}\sigma^2$ 则是有偏的, 又 $E(S^2) = \sigma^2$, 这正是为什么定义样本方差为 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, 而不是 $B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 的原因。

无偏性没有传递性, 如 $E(S^2) = \sigma^2 \nRightarrow E(S) = \sigma$ 。

一般来说, $\bar{X}$ 为总体 X 的 EX 的无偏估计, S^2 为总体 X 的 DX 的无偏估计, 不论什么分布, k 阶样本矩 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ 是总体 k 阶矩的无偏估计。

如果 $E(\hat{\theta}) = a\theta + b \neq \theta$, 即有偏, 则利用下列方法可以修正为无偏估计量:

$$\text{令 } \hat{\theta}_1 = \frac{1}{a}(\hat{\theta} - b), \text{ 则 } E\hat{\theta}_1 = \frac{1}{a}E(\hat{\theta} - b) = \frac{1}{a}(a\theta + b - b) = \theta$$

1.3.2 有效性

在都无偏的条件下, 如 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$ 则 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。

一般来说, μ 的一切形如 $\hat{\mu} = \sum_{i=1}^n a_i X_i$ ($\sum_{i=1}^n a_i = 1$) 的无偏估计中, 当 $a_i = \frac{1}{n}$ 最有效。

1.3.3 一致性 (相合性)

满足 $\lim_{n \rightarrow \infty} P\{|\hat{\theta} - \theta| < \varepsilon\} = 1$ 的 $\hat{\theta}$ 为 θ 的一致估计量, 常使用辛钦大数定理验证。

一般来说, 一致性是自动满足的, 因为点估计的理论根据就是辛钦大数定理。例如: 当用 $\bar{X}$ 估计 μ 时, 根据辛钦大数定理, $\lim_{n \rightarrow \infty} P\{|\bar{X} - \mu| < \varepsilon\} = 1$ 恒成立, 也可以利用切贝雪夫不等式来验证相合性。

一般来说, $\bar{X}$ 为总体 X 的 EX 的一致估计, S^2 为总体 X 的 DX 的一致估计, 若 $\hat{\theta}$ 是 θ 的一致矩估计量, 且 $g(x)$ 在 $x = \theta$ 连续, 则 $g(\hat{\theta})$ 为 $g(\theta)$ 的一致矩估计。若总体 X 的 k 阶原点矩 $EX^k = \mu_k$, 则样本 X 的 k 阶原点矩是总体 X 的 k 阶原点矩 $EX^k = \mu_k$ 一致矩估计; 若 $g(x)$ 为连续函数, 则

$g(A_1, A_2, \dots, A_n)$ 是 $g(\mu_1, \mu_2, \dots, \mu_n)$ 的一致矩估计。

【例】对于正态总体 $N(\mu, \sigma^2)$, 当用 $2\bar{X} - X_1$ 和 $\frac{1}{2}X_1 + \frac{2}{3}X_2 - \frac{1}{6}X_3$ 作为 μ 的估计时, 哪一个更有效?

【解】首先必须验证它们都是无偏估计, 才能通过计算方差大小比较谁更有效。

$$E(2\bar{X} - X_1) = 2E\bar{X} - EX_1 = 2\mu - \mu = \mu$$

$$E\left(\frac{1}{2}X_1 + \frac{2}{3}X_2 - \frac{1}{6}X_3\right) = E\left(\frac{1}{2}X_1\right) + E\left(\frac{2}{3}X_2\right) - E\left(\frac{1}{6}X_3\right) = \frac{1}{2}\mu + \frac{2}{3}\mu - \frac{1}{6}\mu = \mu$$

故它们是无偏估计。

$$D(2\bar{X} - X_1) = D\left(\frac{2}{n} \sum_{i=1}^n X_i - X_1\right) = D\left(\frac{2}{n} \sum_{i=2}^n X_i + \left(\frac{2}{n} - 1\right)X_1\right)$$

$$= \frac{4}{n^2} \sum_{i=2}^n DX_i + \frac{(2-n)^2}{n^2} DX_1 = n^2 \sigma^2 \left(4(n-1) + (2-n)^2\right) = \sigma^2$$

$$D\left(\frac{1}{2}X_1 + \frac{2}{3}X_2 - \frac{1}{6}X_3\right) = \left(\frac{1}{4} + \frac{4}{9} + \frac{1}{36}\right)\sigma^2 = \frac{13}{18}\sigma^2 < \sigma^2$$

故 $\frac{1}{2}X_1 + \frac{2}{3}X_2 - \frac{1}{6}X_3$ 更有效。

【例】证明： S^2 为 σ^2 的一致估计量。

$$\text{【证明】 } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$$

$$ES^2 = \frac{1}{n-1} \left(\sum_{i=1}^n EX_i^2 - nE\bar{X}^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^n (\mu^2 + \sigma^2) - n \left(\mu^2 + \frac{\sigma^2}{n} \right) \right) = \sigma^2; \quad DS^2 = \frac{2\sigma^4}{n-1}$$

根据切比雪夫不等式 $P\{|S^2 - ES^2| < \varepsilon\} \geq 1 - \frac{DS^2}{\varepsilon^2}$

$$\Rightarrow 1 \geq \lim_{n \rightarrow \infty} P\{|S^2 - \sigma^2| < \varepsilon\} \geq \lim_{n \rightarrow \infty} \left(1 - \frac{1}{\varepsilon^2} \cdot \frac{2\sigma^4}{n-1} \right) = 1 \Rightarrow \lim_{n \rightarrow \infty} P\{|S^2 - \sigma^2| < \varepsilon\} = 1, \text{ 结论得证。}$$

【例】设分别来自总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 中抽取容量为 n_1, n_2 的两个独立样本，其样本方差分别为 S_1^2, S_2^2 ，已知 $Y = aS_1^2 + bS_2^2$ 为 σ^2 的最有效估计量，求 a, b 。

【解】已知 $Y = aS_1^2 + bS_2^2$ 为 σ^2 的最有效估计量，首先是无偏估计量，即 $EY = \sigma^2$

$$\Rightarrow aES_1^2 + bES_2^2 = \sigma^2 \Rightarrow a\sigma^2 + b\sigma^2 = \sigma^2 \Rightarrow a + b = 1$$

$$DY = a^2 DS_1^2 + b^2 DS_2^2 = a^2 \cdot \frac{2\sigma^4}{n_1-1} + (1-a)^2 \cdot \frac{2\sigma^4}{n_2-1} = \frac{(n_1+n_2-2)a^2 - 2(n_1-1)a + n_1-1}{(n_1-1)(n_2-1)} 2\sigma^4$$

$$\frac{d}{da}(DY) = 0 \Rightarrow a = \frac{n_1-1}{n_1+n_2-2}, \quad b = \frac{n_2-1}{n_1+n_2-2}$$

【例】设分别来自总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 中抽取容量为 n_1, n_2 的两个独立样本，其样本方差分别为 S_1^2, S_2^2 ，已知 $Y = aS_1^2 + bS_2^2$ 为 σ^2 的最有效估计量，求 a, b 。

【解】已知 $Y = aS_1^2 + bS_2^2$ 为 σ^2 的最有效估计量，首先是无偏估计量，即 $EY = \sigma^2$

$$\Rightarrow aES_1^2 + bES_2^2 = \sigma^2 \Rightarrow a\sigma^2 + b\sigma^2 = \sigma^2 \Rightarrow a + b = 1$$

$$DY = a^2 DS_1^2 + b^2 DS_2^2 = a^2 \cdot \frac{2\sigma^4}{n_1-1} + (1-a)^2 \cdot \frac{2\sigma^4}{n_2-1} = \frac{(n_1+n_2-2)a^2 - 2(n_1-1)a + n_1-1}{(n_1-1)(n_2-1)} 2\sigma^4$$

$$\frac{d}{da}(DY) = 0 \Rightarrow a = \frac{n_1-1}{n_1+n_2-2}, \quad b = \frac{n_2-1}{n_1+n_2-2}$$

【例 16】试讨论正态分布常用一阶矩和二阶矩估计的评价标准。

$$\text{【解】 已知点估计为 } \hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$EX = \mu, ; EX_i = \mu, \quad E\hat{\theta} = E\hat{\mu} = E\bar{X} = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \mu$ ，所以 X_i 和 $\bar{X} = \hat{\mu}$ 都是 μ 的无偏估计量。

$$E(B_2) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E(\bar{X})^2 = \frac{1}{n} \left(\sum_{i=1}^n (\sigma^2 + \mu^2) \right) - \{D(\bar{x}) + E^2(\bar{X})\}$$

$$= \frac{1}{n}(n\sigma^2 + n\mu^2) - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2 \neq \sigma^2. \text{ 所以 } B_2 \text{ 不是 } \sigma^2 \text{ 的无偏估计量;}$$

$$\text{而 } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{nB_2}{n-1} \Rightarrow E(S^2) = E\left(\frac{nB_2}{n-1}\right) = \frac{n}{n-1}E(B_2) = \sigma^2, \text{ 故 } S^2 \text{ 是 } \sigma^2 \text{ 的无偏估计量,}$$

这正是我们称 S^2 为样本方差的理由。

对正态总体 $N(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 和 X_i 都是 $E(X) = \mu$ 的无偏估计量, 但 $D(\bar{X}) = \frac{\sigma^2}{n} \leq D(X_i) = \sigma^2$, 故 $\bar{X}$ 比 X_i 更有效。

【例 17】设总体 X 的分布律为

X	1	2	3
P	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中: $0 < \theta < 1$, $X_1, X_2, \dots, X_n$ 为来自总体 X 的简单随机样本。求

(1) θ 的矩估计 $\hat{\theta}$; (2) θ 的极大似然法估计 $\hat{\theta}$, (3) 判断 θ 的无偏性和一致性。

【解】(1) 根据点估计的定义

$$EX = 1 \times \theta^2 + 2 \times 2\theta(1-\theta) + 3 \times (1-\theta)^2 = 3 - 2\theta \Rightarrow -2\hat{\theta} + 3 = \bar{X} \Rightarrow \hat{\theta} = \frac{3 - \bar{X}}{2}$$

(2) 注意此题一般需要给出样本点来写出联合分布函数。故本题的关键是根据分布率写出代数分布形式, 可以验证: 总体 X 的分布律可以等价写成

$$P\{X=k\} = C_2^{k-1} (1-\theta)^{k-1} \theta^{3-k}, k=1, 2, 3 \quad \left\{ \begin{array}{l} k=1 \Rightarrow C_2^{k-1} (1-\theta)^{k-1} \theta^{3-k} = \theta^2 \\ k=2 \Rightarrow C_2^{k-1} (1-\theta)^{k-1} \theta^{3-k} = 2\theta(1-\theta) \\ k=3 \Rightarrow C_2^{k-1} (1-\theta)^{k-1} \theta^{3-k} = (1-\theta)^2 \end{array} \right.$$

$$\Rightarrow L(x_1, x_2, \dots, x_n, \theta) = P\{X=x_1\} P\{X=x_2\} \cdots P\{X=x_n\}$$

$$= \left[C_2^{x_1-1} (1-\theta)^{x_1-1} \theta^{3-x_1} \right] \left[C_2^{x_2-1} (1-\theta)^{x_2-1} \theta^{3-x_2} \right] \cdots \left[C_2^{x_n-1} (1-\theta)^{x_n-1} \theta^{3-x_n} \right]$$

$$= C_2^{x_1-1} C_2^{x_2-1} \cdots C_2^{x_n-1} (1-\theta)^{\left(\sum_{i=1}^n x_i - n\right)} \theta^{\left(3n - \sum_{i=1}^n x_i\right)}$$

$$\Rightarrow \ln L = \sum_{i=1}^n \ln C_2^{x_i-1} + \left(\sum_{i=1}^n x_i - n\right) \ln(1-\theta) + \left(3n - \sum_{i=1}^n x_i\right) \ln \theta$$

$$\Rightarrow \frac{\partial \ln L}{\partial \theta} = \frac{1}{1-\theta} \left(n - \sum_{i=1}^n x_i\right) + \frac{1}{\theta} \left(3n - \sum_{i=1}^n x_i\right) = 0 \Rightarrow \theta = \frac{3 - \frac{1}{n} \sum_{i=1}^n x_i}{2} \Rightarrow \hat{\theta} = \frac{3 - \frac{1}{n} \sum_{i=1}^n X_i}{2} = \frac{3 - \bar{X}}{2}$$

$$(3) E\hat{\theta} = E\left(\frac{3 - \bar{X}}{2}\right) = \frac{3 - E\bar{X}}{2} = \frac{3 - EX}{2} = \frac{3 - (3 - 2\theta)}{2} = \theta \text{ 故 } \hat{\theta} \text{ 估计是无偏的。}$$

根据大数定理, $\bar{X}$ 依概率收敛到 θ , 于是 $\hat{\theta} = \frac{3 - \bar{X}}{2}$ 依概率收敛到 θ , 即 $\hat{\theta} = \frac{3 - \bar{X}}{2}$ 为 θ 的一致估计量。

二、区间估计 [1]

2.1 置信度与置信区间

点估计是对总体进行了极其“精确”的估计(缩到一点的值),但由于样本的随机性,反而很难接近总体真实值,如果在某个区间估计,将更加合理,所以,人们往往不以得到近似值为满足,还要估计误差,这也是实际工程所需要的结果。

设总体 X 含有一个待估的未知参数 θ , 如果我们从样本 $x_1, x_2, \dots, x_n$ 出发, 找出两个统计量, $\theta_1 = \theta_1(x_1, x_2, \dots, x_n)$ 与 $\theta_2 = \theta_2(x_1, x_2, \dots, x_n)$, $\theta_1 < \theta_2$, 使得区间 (θ_1, θ_2) 以 $1-\alpha$ 的大概率包含这个待估的未知参数 θ , 即 $P\{\theta_1 < \theta < \theta_2\} = 1-\alpha$ ($0 < \alpha < 1$), 那么, 我们称区间 (θ_1, θ_2) 为 θ 的置信区间, $1-\alpha$ 为该区间的置信度(置信水平、置信概率), 上述两个统计量的寻找方法正式源自第六章的 8 个枢轴量。

读者知道, 评价一个点估计的优良有三个标准, 而评价一个区间估计的优劣也有两个因素, 一个是精确度, 即误差大小, 使用区间长度 $L = \theta_2 - \theta_1$ 来刻画, 长度越大, 精确度越低; 第二个因素是可靠性, 使用置信度 $P\{\theta_1 < \theta < \theta_2\} = 1-\alpha$ 来刻画, 置信度 $1-\alpha$ 越大, 即 α 越小, 可靠性越高, 精确度和可靠性是此消彼长的关系, 一般都是在保证满足一定的可靠性(置信度)条件下, 通过增加样本容量, 尽可能提高精确度(即缩小误差)。

2.2 枢轴量

满足以下三个条件的样本函数称为统计量。

- ① 是待估参数 θ 与样本 $\bar{X}$ 的函数;
- ② 不含其它未知参数;
- ③ 服从与未知参数无关的已知分布。

不是统计量就是样本函数, 既可能是样本函数又可能在特定情形下成为统计量的函数称为枢轴量

智轩第 10 技 利用 8 大枢轴量全面解决区间估计和假设检验的 8 大估计方式。

2.3 单正态总体的期望和方差的区间估计

2.3.1 均值 μ 的区间估计

(1) σ^2 已知, 求 μ 的区间估计

选择枢轴量 $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$, 由正态分布图形曲线所围面积可知 $P\left\{\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < z_{\frac{\alpha}{2}}\right\} = 1-\alpha$,

置信度为 $1-\alpha$ 的置信区间为:

$$\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < z_{\frac{\alpha}{2}} \Rightarrow (\theta_1, \theta_2) = \left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \Leftrightarrow \left(\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

事件 $\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < z_{\frac{\alpha}{2}}$ 是根据上分位数定义, 而标准正态分布查表是根据下分位数, 参见浙江大学三版 P61。

如果一律使用使用下分位数,则在求出 θ_1, θ_2 后,把不等式两个端点值交换即可,使用 $z_{\frac{\alpha}{2}}$ 而不使用 z_{α} 的理由是使得 $|\theta_1 - \theta_2|$ 最小,即保证精度尽可能高。注意关系: $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$ 。

(2) σ^2 未知,求 μ 的区间估计

选择枢轴量 $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \Rightarrow P\left\{\left|\frac{\bar{X} - \mu}{S/\sqrt{n}}\right| < t_{\frac{\alpha}{2}}(n-1)\right\} = 1 - \alpha$, 则 $1 - \alpha$ 的置信区间为

$$(\theta_1, \theta_2) = \left(\bar{X} - t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} \right) \Leftrightarrow \left(\bar{X} \pm t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} \right)$$

评注 注意关系: $t_{\frac{\alpha}{2}}(n-1) = -t_{1-\frac{\alpha}{2}}(n-1)$

【例】 已知某种材料的抗压强度 $X \sim N(\mu, \sigma^2)$, 现随机地抽取 16 个试件进行抗压试验, 测得数据如下 (单位: 10^5 Pa):

506 508 499 503 504 510 497 512 514 505 493 496 506 502 509 496.

(1) 求平均抗压强度 μ 的点估计值; (2) 若已知 $\sigma = 6$, 求 μ 的 95% 的置信区间; (3) 求平均抗压强度 μ 的 95% 的置信区间。

【解】 (1) $\hat{\mu} = \bar{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = \frac{1}{16} (506 + 508 + \cdots + 496) = 503.75$ 。

(2) 已知 $\sigma = 6$, 关于 μ 的区间估计应选择随机变量 $U = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$ 。从而, 由 $P\{|U| < z_{\frac{0.05}{2}}\} = 0.95$, 可得置信区间 $(\bar{X} - \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}}, \bar{X} + \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}})$ 。因 $n = 16, z_{0.025} = 1.96$, 故可知 $\frac{\sigma z_{0.025}}{\sqrt{n}} = 6 \times \frac{1.96}{\sqrt{16}} = 2.94$ 。从而所求置信区间为 (500.81, 506.69)。

(3) 由于 σ 未知, 故选用随机变量 $T = \frac{\sqrt{n}(\bar{X} - \mu)}{S}$ 。而相应置信区间为 $\left[\bar{X} \pm \left(\frac{S}{\sqrt{n}} \right) t_{\frac{\alpha}{2}}(n-1) \right]$ 。由样本观测值, 可得 $S = 6.2022$, 且易查表知 $t_{0.025}(15) = 2.1315$ 。于是, 有 $\frac{St_{0.025}(15)}{\sqrt{n}} = 6.2022 \times \frac{2.1315}{\sqrt{16}} = 3.305$ 。因此所求置信区间为 (500.45, 507.05)。

2.3.2 方差 σ^2 的区间估计 ($X \sim N(\mu, \sigma^2)$)

(3) μ 已知, 求 σ^2 的区间估计

选择枢轴量 $\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n) \Rightarrow P\left\{\chi^2_{1-\frac{\alpha}{2}}(n) < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(n)\right\} = 1 - \alpha$, 则 $1 - \alpha$ 的置信区间为

$$(\theta_1, \theta_2) = \left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)} \right)$$

$$\text{或 } S^2 = \frac{n}{n-1} B_2 \Rightarrow (\theta_1, \theta_2) = \left(\frac{nB_2}{\chi_{\frac{\alpha}{2}}^2(n)}, \frac{nB_2}{\chi_{1-\frac{\alpha}{2}}^2(n)} \right) \quad \text{其中 } B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

(4) μ 未知, 求 σ^2 的区间估计

$$\text{取枢轴量 } \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \Rightarrow P \left\{ \chi_{1-\frac{\alpha}{2}}^2(n-1) < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} < \chi_{\frac{\alpha}{2}}^2(n-1) \right\} = 1-\alpha$$

$$(\theta_1, \theta_2) = \left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \right)$$

【例】 设某种清漆的 9 个样品, 其干燥时间 (单位: h) 分别为 6.0, 5.7, 5.8, 6.5, 7.0, 6.3, 5.6, 6.1, 5.0; 设干燥时间总体服从正态分布 $N(\mu, \sigma^2)$ 。分别就下列两种情形求 σ^2 的置信度为 0.95 的置信区间: (1) 若由以往经验知 $\mu = 6$; (2) μ 为未知。

【解】(1) 由于 μ 已知, 确定 σ^2 的置信区间, 应选取函数 $\chi^2 = [\sum_{i=1}^n (X_i - \mu)^2] / \sigma^2$, 且置信度为 $1-\alpha$ 的置信区间为 $([\sum_{i=1}^n (X_i - \mu)^2] / \chi_{\frac{\alpha}{2}}^2(n), [\sum_{i=1}^n (X_i - \mu)^2] / \chi_{1-\frac{\alpha}{2}}^2(n))$ 。

由样本值 $x_i (i=1, 2, \dots, 9)$ 及 $\mu = 6, n = 9$, 可计算得 $\sum_{i=1}^9 (x_i - 6)^2 = 3.04$ 。

又由 $1-\alpha = 0.95$, 知 $\frac{\alpha}{2} = 0.025$; 从而查 χ^2 分布表, 可得 $\chi_{0.025}^2(9) = 19.023$, $\chi_{0.975}^2(9) = 2.700$ 。

因此所求置信区间为 $(\frac{3.04}{19.023}, \frac{3.04}{2.700}) = (0.160, 1.126)$ 。

(2) 当 μ 未知时, 确定 σ^2 的置信区间, 应选取随机变量 $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$, 此时, 置信度为 $1-\alpha$ 的 σ^2 的置信区间为 $(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)})$ 。

由样本值, 可计算得 $S^2 = 0.577^2$; 又查表可知 $\chi_{0.025}^2(8) = 17.535$, $\chi_{0.975}^2(8) = 2.180$ 。故所求置信区间

为 $(8 \times \frac{0.577^2}{17.535}, 8 \times \frac{0.577^2}{2.180}) = (0.152, 1.223)$ 。

2.4 两个正态总体均值差与方差比的区间估计

$$X \sim N(\mu, \sigma_1^2); Y \sim N(\mu_2, \sigma_2^2); \quad X \text{ 和 } Y \text{ 独立同}, \quad \bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}).$$

2.4.1 $\mu_1 - \mu_2$ 区间估计(5) σ_1^2, σ_2^2 已知, 区间估计 $\mu_1 - \mu_2$

$$\text{取枢轴量 } U = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1) \Rightarrow P\left\{\left|\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right| < z_{\frac{\alpha}{2}}\right\} = 1 - \alpha, \text{ 置信度为 } 1 - \alpha \text{ 的置信区}$$

$$\text{间为 } (\theta_1, \theta_2) = \left(\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}, \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right) \Leftrightarrow \left(\bar{X} - \bar{Y} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right)$$

【例】设两总体 X, Y 相互独立, $X \sim N(\mu_1, 64), Y \sim N(\mu_2, 36)$, 从 X 中取出 $n_1 = 75$ 的样本, 求得 $\bar{x} = 82$;

从 Y 中取出 $n_2 = 50$ 的样本, 求得 $\bar{y} = 76$ 。试求 $(\mu_1 - \mu_2)$ 的 96% 的置信区间。

【解】这是两正态总体在 σ_1^2, σ_2^2 已知的情况下, 对期望差 $(\mu_1 - \mu_2)$ 作区间估计。置信度为 $1 - \alpha$ 的置信区

间为 $[(\bar{x} - \bar{y}) - \delta, (\bar{x} - \bar{y}) + \delta]$, 其中 $\delta = z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ 。由于 $1 - \alpha = 0.96$,

$$\alpha = 0.04, \frac{\alpha}{2} = 0.02; z_{\frac{\alpha}{2}} = z_{0.02} \approx 2.054, \delta = 2.054 \times \sqrt{\frac{64}{75} + \frac{36}{50}} \approx 2.58; \bar{x} - \bar{y} = 82 - 76 = 6.$$

故 $\mu_1 - \mu_2$ 的置信区间为 $(6 - 2.58, 6 + 2.58) = (3.42, 8.58)$ 。

(6) σ_1^2, σ_2^2 未知但 $\sigma_1^2 = \sigma_2^2 = \sigma^2$, 估计 $\mu_1 - \mu_2$

$$\text{取枢轴量 } T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2), \text{ 其中 } S_w^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n + m - 2},$$

$$\Rightarrow P\left\{\left|\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{m} + \frac{1}{n}}}\right| < t_{\frac{\alpha}{2}}(n + m - 2)\right\} = 1 - \alpha, \text{ 则置信度为 } 1 - \alpha \text{ 的置信区间为}$$

$$(\theta_1, \theta_2) = \left(\bar{X} - \bar{Y} \pm t_{\frac{\alpha}{2}}(n) \cdot (n + m - 2) S_w \sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

【例】为了估计磷肥对某种农作物增产的作用, 现选 30 块条件大致相同的土地。10 块不施磷肥, 另外 10 块施磷肥, 得亩产量 (单位: kg) 如下: 不施磷肥亩产 560, 590, 560, 570, 580, 570, 600, 550, 570, 550; 施磷肥亩产 620, 570, 650, 600, 630, 580, 570, 600, 600, 580。设亩产服从正态分布, 且方差相同; 取置信度为 0.95, 试对施磷肥平均亩产和不施磷肥平均亩产之差作区间估计。

〔解〕 把不施磷肥亩产看成总体 $X \sim N(\mu_1, \sigma_1^2)$, 施磷肥亩产看成总体 $Y \sim N(\mu_2, \sigma_2^2)$, $\sigma_1^2 = \sigma_2^2$ 。本例

则为求 $\mu_2 - \mu_1$ 的区间估计问题。由题设, 知 $n_1 = n_2 = 10$, 经计算可得:

$$\bar{x} = 570, n_1 s_{1n_1}^2 = \sum_{i=1}^{10} (x_i - \bar{x})^2 = 2400, \bar{y} = 600, n_2 s_{2n_2}^2 = \sum_{i=1}^{10} (y_i - \bar{y})^2 = 6400。$$

又由 $1 - \alpha = 0.95$, 查 t 分布表, 得 $t_{0.025}(18) = 2.1009$ 。因此 $\mu_2 - \mu_1$ 的置信下限为

$$\bar{y} - \bar{x} - t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) \sqrt{(n_1 s_{1n_1}^2 + n_2 s_{2n_2}^2)(n_1 + n_2) / n_1 n_2 (n_1 + n_2 - 2)}$$

$$= 600 - 570 - 2.1009 \times \sqrt{\frac{(2400 + 6400) \times 20}{10 \times 10 \times 18}} = 9;$$

而置信上限为

$$\bar{y} - \bar{x} + t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) \sqrt{(n_1 s_{1n_1}^2 + n_2 s_{2n_2}^2)(n_1 + n_2) / n_1 n_2 (n_1 + n_2 - 2)}$$

$$= 600 - 570 + 2.1009 \times \sqrt{\frac{(2400 + 6400) \times 20}{10 \times 10 \times 18}} = 51。$$

2.4.2 σ_1^2 / σ_2^2 的区间估计

(7) μ_1, μ_2 已知, 估计方差比 $\frac{\sigma_1^2}{\sigma_2^2}$ 的置信区间

$$\text{选择枢轴量 } Q = \frac{\frac{1}{m\sigma_1^2} \sum_{i=1}^m (X_i - \mu_1)^2}{\frac{1}{n\sigma_2^2} \sum_{i=1}^n (Y_i - \mu_2)^2} \sim F(m, n) \Rightarrow P \left\{ F_{1-\frac{\alpha}{2}}(m, n) < \frac{\frac{1}{m\sigma_1^2} \sum_{i=1}^m (X_i - \mu_1)^2}{\frac{1}{n\sigma_2^2} \sum_{i=1}^n (Y_i - \mu_2)^2} < F_{\frac{\alpha}{2}}(m, n) \right\} = 1 - \alpha$$

则置信度为 $1 - \alpha$ 的置信区间为

$$(\theta_1, \theta_2) = \left(\frac{\frac{1}{m} \sum_{i=1}^m (X_i - \mu_1)^2}{\frac{1}{n} \sum_{i=1}^n (Y_i - \mu_2)^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(m, n)}, \frac{\frac{1}{m} \sum_{i=1}^m (X_i - \mu_1)^2}{\frac{1}{n} \sum_{i=1}^n (Y_i - \mu_2)^2} \cdot \frac{1}{F_{1-\frac{\alpha}{2}}(m, n)} \right)$$

(8) μ_1, μ_2 未知, 估计方差比 $\frac{\sigma_1^2}{\sigma_2^2}$ 的置信区间

$$\text{选择枢轴量 } Q = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(m-1, n-1) \Rightarrow P \left\{ F_{1-\frac{\alpha}{2}}(m-1, n-1) < \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} < F_{\frac{\alpha}{2}}(m-1, n-1) \right\} = 1 - \alpha,$$

则置信度为 $1 - \alpha$ 的置信区间为

$$(\theta_1, \theta_2) = \left(\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(m-1, n-1)}, \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{\alpha}{2}}(m-1, n-1)} \right)$$

$$\text{或 } (\theta_1, \theta_2) = \left(F_{1-\frac{\alpha}{2}}(n-1, m-1) \frac{S_1^2}{S_2^2}, F_{\frac{\alpha}{2}}(n-1, m-1) \frac{S_1^2}{S_2^2} \right),$$

评注 注意关系: $F_{1-\frac{\alpha}{2}}(m-1, n-1) = \frac{1}{F_{\frac{\alpha}{2}}(n-1, m-1)}$; $F(m, n) = \frac{1}{F(n, m)}$ 。

注意, 考研范围仅限于上述 8 个正态整体的双侧置信区间, 其它双侧不做要求, 读者不要作无用功。

【例】 两总体 X, Y 相互独立, $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 分别取 $n_1 = 25, n_2 = 16$ 的简单随机样本, 算出 $s_1^2 = 63.96, s_2^2 = 49.05$ 。试求两总体方差比 $\frac{\sigma_1^2}{\sigma_2^2}$ 的 98% 的置信区间。

〔解〕这是两正态总体在 μ_1, μ_2 都未知的情况下, 求方差比 $\frac{\sigma_1^2}{\sigma_2^2}$ 置信度为 $1-\alpha$ 的置信区间:

$$\left(\frac{s_1^2 / s_2^2}{F_{\frac{\alpha}{2}}(n_1-1, n_2-1)}, \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}}(n_2-1, n_1-1) \right)。$$

因为 $1-\alpha = 0.98$, 故 $\alpha = 0.02, \frac{\alpha}{2} = 0.01$; 又因 $n_1 - 1 = 24, n_2 - 1 = 15$, 所以, 有

$$F_{0.01}(24, 15) = 3.29, F_{0.01}(15, 24) = 2.89。$$

从而, 知 $\frac{s_1^2 / s_2^2}{F_{\frac{\alpha}{2}}(n_1-1, n_2-1)} = \frac{63.96}{49.05} \times \frac{1}{3.29} \approx 0.396, \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}}(n_2-1, n_1-1) = \frac{63.96}{49.05} \times 2.89 \approx 3.768$, 因而

(0.396, 3.768) 为所求。

2.5 单侧置信区间

定义: $P\{\theta > \theta_L\} = 1-\alpha$ 或 $P\{\theta < \theta_u\} = 1-\alpha$, 则 $(\theta_L, +\infty), (-\infty, \theta_u)$ 为 $1-\alpha$ 的单侧置信区间。

确定技巧: 利用前述的双侧区间的上下限, 把 $\frac{\alpha}{2}$ 改为 α , 就可以了。如

a) $X \sim N(\mu, \sigma^2), \mu, \sigma^2$ 均未知, 估计 μ (单侧) 下上侧置信区间为

$$(\theta_L, +\infty) = \left(\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha}(n-1), +\infty \right) \text{ 或 } (-\infty, \theta_u) = \left(-\infty, \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha}(n-1) \right)$$

b) $X \sim N(\mu, \sigma^2), \mu, \sigma^2$ 均未知, 单侧估计 σ^2

$$\sigma^2 \text{ 的置信度为 } 1-\alpha \text{ 的单侧置信区间为 } \left(0, \frac{(n-1)s^2}{x_{1-\alpha}^2(n-1)} \right)$$

注意, 考研范围仅限于 4 个单正态整体的单侧置信区间。

第八章 假设检验 [数学 I]

本章导读 假设检验 8 大接受域和拒绝域, 两类错误的计算方法。

一、假设检验与参数区间估计的关系

1. 参数 θ 的置信度为 $1-\alpha$ 的区间估计, 正好是显著性水平为 α 的假设检验的接受域。
2. 区间估计中, 假设总体中的参数是未知的, 要用样本对它进行估计; 而假设检验中, 是先对参数做出假设, 再用样本值对假设作检验。在某种意义上, **假设检验是区间估计的逆问题**。
3. 具有**完全相同的 8 大枢轴量** (8 大枢轴量详见第六章)。

二、假设检验的基本思想及两类错误与显著性检验

假设检验的基本思想是根据**小概率原理**, 即为概率很小的事件在一次试验中是几乎不可能发生的。首先, 提出假设, 并在假设成立的条件下, 由**一次抽样的样本值**计算出事件 A 是小概率事件, 于是使我们不得不怀疑假设的合理性, 从而否定假设; 否则就可以认为假设合理。

比如, 一个人说他射击是高手, 我们将半信半疑。怎样才能确定他的话真假, 最好的办法就是先假设他是高手或低手, 然后让他实际打几枪, 根据他这次射击的结果来检验。如果其射击结果命中率在 90% 以上, 我们一般可以接受他的说法; 如果命中率在 50% 以下, 我们就拒绝他的说法, 这个判断的标准是根据小概率事件几乎不可能发生的原理。但我们的判断也可能犯错误, 因为几乎不可能发生事件, 毕竟还是有可能发生的。一是他的确是高手, 但在这次射击中失误了, 而我们却只根据他这一次的命中率没把他当高手, 也就是说我们犯了以真当假的错误—称为**第一类错误**。二是他本来是个低手, 但这次命中率恰好超过了 90% 以上, 我们却把他当成了高手, 事实上我们犯了以假当真的错误—称为**第二类错误**。这两类错误, 我们都尽可能使其概率最小, 但事实上做不到, 因为它们是如此消彼长的关系, 因此, 我们首先要控制主要错误 (又称显著性错误) 的概率, 那究竟哪种错误严重, 即显著呢?

为了说明两类错误主次关系的直观含义, 我们引用一个生活例子: 某人因身体不适前往医院就医。医生的职责就是通过各种生理检查, 根据化验的数据作出该病员是否犯病的结论。然而再好的医生都不可避免会犯下两类错误。一种是病员确实有病, 但由于生理指标未出现明显的异常现象, 使医生判断为无病。另一种是病员实际上没有疾病, 但生理指标呈现某种异常, 使医生判断为有病。这两类错误都会导致病员的损失, 然而两类错误的损失是不一样的。如果“有病判无病”, 相当于以真当假—第 1 类错误, 其结果可能延误了治病的时机造成病情的加重以致死亡; 而“无病判有病”, 相当于以假当真—第 2 类错误, 其结果是病员会有一些经济或其他损失, 然而对生命是无碍的。因此医生总是尽可能地避免犯上述第 1 类错误。

如果用假设检验的数学语言表达即为: 原假设 H_0 : 该人有病, 备选假设 H_1 : 该人无病。而生理检验指标就是样本观测值, 医生要作的决定即为接受 H_0 或拒绝 H_0 。而拒绝域 W 即为“生理指标属于正常范围”, 而犯第 I 类错误的概率为 $P_{H_0}(\text{拒绝 } H_0 | H_0 \text{ 为真}) = \alpha$, 正是“有病误判无病”的概率。犯第 II 类错误的概率为 $P_{H_1}(\text{接受 } H_0 | H_1 \text{ 为真}) = P(\text{接受 } H_0 | H_0 \text{ 不真}) = \beta$, 正是“无病误判有病”的概率。由于两类错误的影响不同, 第一类错误的后果严重, 是显著的, 故经常要求首先控制第一类错误概率, 即 $P_{H_0}(W) \leq \alpha$, 其中 α 是充分小的正数, 这也就是**显著性检验**的由来, α 称为**显著性水平**。 α 的意思是概率

P_{H_0} (拒绝 H_0 | H_0 为真) 很小, 相应的事件为小概率事件, 几乎不可能发生, 但并不等于不发生, 只是发生的概率等于 α 而已。注意显著性检验并不考虑犯第 2 类错误概率 β 。

当总体的分布函数未知, 或只知其形式而不知道它的若干参数的情况时, 我们可以根据历史经验或者其他的充分理由对总体分布或者它的参数提出某种假设: 例如

H_0 : 总体分布是正态分布 或者 $H_0: \theta_1 \leq \theta_{10}$ (θ_{10} 为某一已知常数), 称为原假设;

而它的否定命题为: H_1 : 总体分布不是正态分布 或者 $H_1: \theta_1 > \theta_{10}$, 称为备择假设。

假设检验方案即为寻求一个拒绝区间 W , 一旦样本的观测值落在 W 内, 就作出拒绝 H_0 的决定, 故称为拒绝域, W 的余集 $\bar{W}$ 成为接受域, 接受域正是区间估计的范围。样本的观测值落在 $\bar{W}$ 就作出接受 H_0 的决定。假设检验就是根据取到的样本观测值计算得到的枢轴量检验值与根据查表得到的枢轴量理论值作比较, 作出拒绝还是接受哪个假设的决定。

例如若 $P\left\{\left|\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\right| \geq \lambda\right\} = \alpha$, 当 $\alpha = 0.01; 0.05, 0.1$ 等很小的值时就是小概率事件, 小概率事件

在一次试验中认为是不会发生的, 也就是说事件 $\left\{\left|\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\right| \geq \lambda\right\}$ 是不可能发生的, 即 $\left|\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\right| < \lambda$ 关系成立,

则原假设正确; 如果发生了, 即 $\left|\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\right| \geq \lambda$ 关系成立, 则原假设错误。

例如, 已知方差 σ_0 和 α , 在假设 $H_0: \mu = \mu_0; H_1: \mu \neq \mu_0$ 下, 由于 $\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$, 其中分位数

$z = z_{\frac{\alpha}{2}}$ 或 $z = z_{1-\frac{\alpha}{2}}$ 可查表得到理论值, 即得 λ 值, $\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}$ 的检验值可由样本观察值和 $\mu = \mu_0$ 计算确定。

如果将样本观察值 $\bar{X}$ 和原假设 H_0 中的 $\mu = \mu_0$ 代入 $\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}$, 有 $\left|\frac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}}\right| \geq z_{1-\frac{\alpha}{2}}$ (或 $\left|\frac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}}\right| \geq z_{\frac{\alpha}{2}}$ 成立,

到底取哪一种形式, 由取上分位数还是下分位数决定, 依据是要看哪一种能查表得出) 则小概率事件发生了, 就拒绝 H_0 ; 反之, 接受 H_0 。显然,

$\left|\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\right| \geq z_{1-\frac{\alpha}{2}}$ 的拒绝域为 $\left(-\infty, -z_{1-\frac{\alpha}{2}}\right) \cup \left(z_{1-\frac{\alpha}{2}}, +\infty\right)$; 接受域

为 $\left(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)$, 由此还可以推出 μ 的拒绝域为 $\mu \geq \bar{X} + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \cup \mu \leq \bar{X} - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$ 。可以看出, 它与区间估计的分界点是完全相同的, 而区域描述正好互补。

当假设形式为: $H_0: \mu = \mu_0; H_1: \mu \neq \mu_0$ 时, 只要比较 Z_0 (枢轴量) 的检验计算值与分位数的查表值 $z_{\frac{\alpha}{2}}$ 的大小即可, 属于双边检验, 即有结论:

(a) $|Z_0| > z_{\frac{\alpha}{2}} \Rightarrow Z_0$ 落在拒绝域内, 从而拒绝原假设 H_0 , 认为 H_1 正确;

(b) $|Z_0| \leq z_{\frac{\alpha}{2}} \Rightarrow Z_0$ 未落在拒绝域内, 从而接受原假设 H_0 , 认为 H_0 正确。

但是, 如果假设形式为在假设 $H_0: \mu = \mu_0; H_1: \mu > \mu_0$ 或 $H_0: \mu = \mu_0; H_1: \mu < \mu_0$ 下, 就要比较 Z_0 (枢

轴量)的检验计算值与分位数的查表值 Z_α 的大小,属于单边检验,即:

(a)' $Z_0 > z_\alpha \Rightarrow Z_0$ 落在拒绝域内,从而拒绝原假设 H_0 , 认为 H_1 正确;

(b)' $Z_0 \leq z_\alpha \Rightarrow Z_0$ 未落在拒绝域内,从而接受原假设 H_0 , 认为 H_0 正确。

其余枢轴量类推,参见下表。

三、正态总体的参数检验的考点

3.1 单个正态总体

注意等号的分布位置,对单边假设等号一般在原假设上。

检验参数	条件	H_0	H_1	H_0 的拒绝域	选用检验枢轴量	自由度	分位点
数学期望 μ	σ^2 已知	$\mu = \mu_0$	$\mu \neq \mu_0$	$ Z \geq Z_{\alpha/2} = -Z_{1-\alpha/2}$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ $\sim N(0,1)$		$\pm Z_{\alpha/2}$
		$\mu \leq \mu_0$	$\mu > \mu_0$	$Z \geq Z_\alpha = -Z_{1-\alpha}$			Z_α
		$\mu \geq \mu_0$	$\mu < \mu_0$	$Z \leq -Z_\alpha = Z_{1-\alpha}$			$-Z_\alpha$
	σ^2 未知	$\mu = \mu_0$	$\mu \neq \mu_0$	$ t \geq t_{\alpha/2} = -t_{1-\alpha/2}$	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ $\sim t(n-1)$	$n-1$	$\pm t_{\alpha/2}$
		$\mu \leq \mu_0$	$\mu > \mu_0$	$t \geq t_\alpha = -t_{1-\alpha}$			t_α
		$\mu \geq \mu_0$	$\mu < \mu_0$	$t \leq -t_\alpha = t_{1-\alpha}$			$-t_\alpha$
方差 σ^2	μ 已知	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\begin{cases} \chi^2 \geq \chi_{\alpha/2}^2 \\ \text{或} \chi^2 \leq \chi_{1-\alpha/2}^2 \end{cases}$	$\chi^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2}$ $\sim \chi^2(n)$	n	$\begin{cases} \chi_{\alpha/2}^2 \\ \chi_{1-\alpha/2}^2 \end{cases}$
		$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi_\alpha^2$			χ_α^2
		$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 \leq \chi_{1-\alpha}^2$			$\chi_{1-\alpha}^2$
	μ 未知	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\begin{cases} \chi^2 \geq \chi_{\alpha/2}^2 \\ \text{或} \chi^2 \leq \chi_{1-\alpha/2}^2 \end{cases}$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ $\sim \chi^2(n-1)$	$n-1$	$\begin{cases} \chi_{\alpha/2}^2 \\ \chi_{1-\alpha/2}^2 \end{cases}$
		$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi_\alpha^2$			χ_α^2
		$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2 \leq \chi_{1-\alpha}^2$			$\chi_{1-\alpha}^2$

注 对同一问题进行双侧与单侧检验,可能得出不同的结论。这是因为尽管统计量的观测值不变,但由于双侧与单侧检验有着不同的拒绝域,从而同一统计量的值可能落入不同的拒绝域(或接受域)。所以可能得到不同的结论(一般,单侧检验的拒绝域包含双侧检验中的对应一侧的拒绝域(因为 α 越大,拒绝域越大(犯第一类错误的概率亦越大))。从而,往往在双侧检验中被接受的结论,可能在单侧检验中却被拒绝)。

3.2 两个正态总体的假设检验

(1) 已知方差 σ_1^2 和 σ_2^2 , 检验均值 μ (Z 检验)

原假设 H_0	备样假设 H_1	H_0 成立时统计量	α 显著性水平下 H_0 的拒绝域
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$ $\sim N(0,1)$	$ U \geq Z_{\alpha/2}$
	$\mu_1 < \mu_2$		$U \geq Z_{\alpha}$
	$\mu_1 > \mu_2$		$U \leq -Z_{\alpha}$

(2) 未知 σ_1^2 及 σ_2^2 , 但 $\sigma_1^2 = \sigma_2^2$, (其中 $S_w^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$), 检验均值 (t 检验)

原假设 H_0	备样假设 H_1	H_0 成立时的统计量	α 显著性水平下 H_0 的拒绝域
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$T = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{\frac{1}{m} + \frac{1}{n}}}$ $\sim t(m+n-2)$	$ T \geq t_{\frac{\alpha}{2}}(m+n-2)$
	$\mu_1 < \mu_2$		$T \geq t_{\alpha}(m+n-2)$
	$\mu_1 > \mu_2$		$T \leq -t_{\alpha}(m+n-2) = t_{1-\alpha}$

(3) 已知均值 μ_1 与 μ_2 , 检验方差 (F 检验) (注意: $F_{1-\alpha}(n, m) = \frac{1}{F_{\alpha}(n, m)}$)

原假设 H_0	备样假设 H_1	H_0 成立时的统计量	α 显著性水平下 H_0 的拒绝域
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{\frac{1}{m} \sum_{i=1}^m (x_i - \mu_1)^2}{\frac{1}{n} \sum_{j=1}^n (y_j - \mu_2)^2} \sim F(m, n)$	$F \geq F_{\frac{\alpha}{2}}(m, n)$ 或 $F \leq F_{1-\frac{\alpha}{2}}(m, n)$
	$\sigma_1^2 < \sigma_2^2$		$F \geq F_{\alpha}(m, n)$
	$\sigma_1^2 > \sigma_2^2$		$F \leq F_{1-\alpha}(m, n)$

(4) 未知均值 μ_1 和 μ_2 , 检验方差 (F 检验)

原假设 H_0	备样假设 H_1	H_0 成立时的统计量	α 显著性水平下 H_0 的拒绝域
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$ $\sim F(m-1, n-1)$	$F \geq F_{\frac{\alpha}{2}}(m-1, n-1)$ 或 $F \leq F_{1-\frac{\alpha}{2}}(m-1, n-1)$
	$\sigma_1^2 < \sigma_2^2$		$F \geq F_{\alpha}(m-1, n-1)$
	$\sigma_1^2 > \sigma_2^2$		$F \leq F_{1-\frac{\alpha}{2}}(m-1, n-1)$

$$\text{其中: } S_1 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \mu_1)^2 \quad S_2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_2)$$

四、假设检验的符号设定与计算技巧

4.1 假设检验的符号法则

假设检验的假设包括原假设 H_0 和备择假设 H_1 。原假设 H_0 中一般都要含有“=”，可以是“=”，“≥”，“≤”，但不能是“≠”，“>”或“<”。因为在进行假设检验计算时要将原假设代入统计量，若原假设不含有“=”，则无法将之代入统计量进行计算，这就是原假设 H_0 中一定要含有“=”的原因。

在假设检验中， $H_0: \mu = \mu_0$ ； $H_1: \mu > \mu_0$ 与 $H_0: \mu \leq \mu_0$ ； $H_1: \mu > \mu_0$ 是等价的； $H_0: \mu = \mu_0$ ； $H_1: \mu < \mu_0$ 与 $H_0: \mu \geq \mu_0$ ； $H_1: \mu < \mu_0$ 也是一个意思。

比如要检验参数 μ ，根据不同要求，可以做原假设 $H_0: \mu = \mu_0$ ，也可以做原假设 $H_0: \mu \leq \mu_0$ ，或者可以做 $H_0: \mu \geq \mu_0$ 。但不能作原假设 $H_0: \mu \neq \mu_0$ ，也不能做 $H_0: \mu > \mu_0$ 或 $H_0: \mu < \mu_0$ 。

若原假设为 $H_0: \mu = \mu_0$ ，则备择假设可以是 $H_1: \mu \neq \mu_0$ ， $H_1: \mu > \mu_0$ 或 $H_1: \mu < \mu_0$ 三种情形都可以；若原假设是 $H_0: \mu \leq \mu_0$ ，则备择假设通常是 $H_1: \mu > \mu_0$ ；若原假设是 $H_0: \mu \geq \mu_0$ ，则备择假设通常是 $H_1: \mu < \mu_0$ 。

4.2 统计量的接受域与拒绝域快速求法及一般步骤

4.2.1 接受域与拒绝域的“位置关系”

设 μ_0 为某一固定值，若 $\mu < \mu_0$ 成立，则称 μ 在 μ_0 之左；若 $\mu > \mu_0$ 成立，则称 μ 在 μ_0 之右；若 $\mu \neq \mu_0$ 成立，则称 μ 在 μ_0 之两侧。这可以通过数轴上点的位置关系画图看出。

根据原假设和备择假设的不同形式，可以规定 H_0 ， H_1 的位置关系：

若 $H_0: \mu \leq \mu_0$ (或 $H_0: \mu = \mu_0$)， $H_1: \mu > \mu_0$ ，称 H_0 在左， H_1 在右。

若 $H_0: \mu \geq \mu_0$ (或 $H_0: \mu = \mu_0$)， $H_1: \mu < \mu_0$ ，称 H_0 在右， H_1 在左。

若 $H_0: \mu = \mu_0$ ， $H_1: \mu \neq \mu_0$ ，称 H_0 在中间， H_1 在两侧。

H_1 相对 H_0 的位置关系就是 H_1 中 μ 相对于已知量 μ_0 的位置关系。

设显著性水平为 α ，对于不同类型的假设检验，可以用下面的方法确定“统计量”的接受域与拒绝域（一般只写拒绝域）：

在 8 个枢轴量中，根据题意找出统计 X ，设统计量服从的分布为 A （ A 可以是正态分布、 t 分布、 χ^2 分布、 F 分布）。判断 H_0 与 H_1 的位置关系， H_0 对应“统计量”的接受域， H_1 对应“统计量”的拒绝域。

通过 H_0 与 H_1 的位置关系形象地确定接受域与拒绝域的位置关系。接受域较大，对应概率为 $1-\alpha$ ，拒绝域较小，对应概率为 α ，这就可以形象地确定分位点，从而得到各种类型的“统计量”的接受域与拒绝域。

若 A 的概率密度关于 y 轴对称，即为偶函数时（正态分布， t 分布），根据对称性，接受域可以写为

$|X| < z_{1-\frac{\alpha}{2}}$ 或 $|X| < z_{\frac{\alpha}{2}}$, 拒绝域可以写为 $|X| > z_{1-\frac{\alpha}{2}}$ 或 $|X| > z_{\frac{\alpha}{2}}$ 。

明确了“统计量”的接受域与拒绝域的概念, 就可以进一步得到假设检验的一般步骤。

4.2.2 假设检验计算的一般步骤如下

根据上面对“统计量”的接受域与拒绝域的讨论, 可以进一步得到假设检验的一般步骤:

- (1) 根据已知条件, 确定原假设和备择假设, 并找出统计量。
- (2) 判断 H_0 , H_1 的位置关系。根据 H_0 对应概率为 $1-\alpha$ 的接受域, H_1 对应概率为 α 的拒绝域的原则, 并结合它们的位置关系确定分位点, 再利用分位点写出“统计量”的接受域或拒绝域 (一般只写拒绝域即可)。将 H_0 中的等式 $\mu = \mu_0$ 和样本值代入统计量进行计算, 若结果落在“统计量”的接受域, 就保留其对应的原假设 H_0 ; 若结果落在“统计量”的拒绝域, 就保留其对应的备择假设 H_1 。

五、两类错误的概率关系与计算方法

参数的假设检验是利用一个样本值来推断总体, 由于样本的随机性, 不可避免地会发生错误: “弃真错误—第一类错误”和“取伪错误—第二类错误”。两类错误的计算主要目的就是反求 α 。

1. 第一类错误的概率定义 $\alpha = P\{\text{拒绝}H_0 | H_0\text{真}\} = P\{\text{拒绝}H_0 | H_1\text{假}\}$

2. 第二类错误的概率定义 $\beta = P\{\text{接受}H_0 | H_1\text{真}\} = P\{\text{接受}H_0 | H_0\text{假}\}$

设正态总体的方差 $\sigma^2 = \sigma_0^2$, μ 只可能取 μ_0 或 $\mu_1 (> \mu_0)$ 二者之一, 又设 $X_1, X_2, \dots, X_n$ 是来自此总体的简单随机样本, 检验假设 $H_0: \mu = \mu_0$; $H_1: \mu = \mu_1 (> \mu_0)$, 则两类错误的概率计算如下

因样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 是总体均值 μ 的无偏估计量且 $\mu_0 < \mu_1$, 因此当 $\bar{X} - \mu_0$ 过分偏大时, 则说明

H_0 不真, 故拒绝域的形式为 $\bar{X} - \mu_0 \geq c$, 故

犯第一类错误的概率为

$$\begin{aligned}\alpha &= P\{\text{拒绝}H_0 | H_0\text{为真}\} = P\{\bar{X} - \mu_0 \geq c | H_0\text{为真}\} = P\{\bar{X} - \mu_0 \geq c | \mu = \mu_0\} \\ &= P\left\{\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \geq \frac{c\sqrt{n}}{\sigma_0} \mid \mu = \mu_0\right\} = P\left\{\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} \geq \frac{c\sqrt{n}}{\sigma_0}\right\} = 1 - P\left\{\frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} < \frac{c\sqrt{n}}{\sigma_0}\right\} = 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma_0}\right) = 1 - \Phi(u_\alpha)\end{aligned}$$

$$\text{其中 } u_\alpha = \frac{c\sqrt{n}}{\sigma_0}.$$

犯第二类错误的概率为

$$\begin{aligned}\beta &= P\{\text{接受}H_0 | H_1\text{为真}\} = P\{\bar{X} - \mu_0 < c | H_1\text{为真}\} = P\{\bar{X} < c + \mu_0 | \mu = \mu_1\} \\ &= P\left\{\frac{\bar{X} - \mu_1}{\sigma_0/\sqrt{n}} \geq \frac{(c + \mu_0 - \mu_1)\sqrt{n}}{\sigma_0} \mid \mu = \mu_1\right\} = P\left\{\frac{\bar{X} - \mu_1}{\sigma_0/\sqrt{n}} < \frac{(c + \mu_0 - \mu_1)\sqrt{n}}{\sigma_0} \mid \mu = \mu_1\right\} = P\left\{\frac{\bar{X} - \mu_1}{\sigma_0/\sqrt{n}} < \frac{[c - (\mu_1 - \mu_0)]\sqrt{n}}{\sigma_0}\right\} \\ &= \Phi\left(\frac{c\sqrt{n}}{\sigma_0} - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right) = \Phi\left(u_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0/\sqrt{n}}\right)\end{aligned}$$

$\Phi(x)$ 单调不减, 如果减少 α , 则 u_α 增大, 故 $u_\alpha - \frac{\mu_1 - \mu_0}{\sigma_0 / \sqrt{n}}$ 也增大, 从而 β 也增大, 同理, α 增加则 β 也

减少。可见, 同时减少 α, β 是不可能的, 但我们可以适当调整 c 的值, 使得在首先控制 α 的情形下, 使得 β 尽可能小, 这正是显著性检验的思想。

为了让读者彻底搞清楚两类错误的关系和计算方法, 我们从另外一个角度来阐述: 设总体 $X \sim N(\mu, \sigma^2)$, 方差 σ^2 已知, 关于均值 μ 的双侧假设检验, 给出计算两类错误的公式如下

设 $\sigma^2 = \sigma_0^2$, 检验假设 $H_0: \mu = \mu_0; H_1: \mu = \mu_1 \neq \mu_0$, 则 H_0 的接受域为

$$(\lambda_1, \lambda_2) = \left(\mu_0 - \frac{z_{\frac{\alpha}{2}} \cdot \sigma_0}{\sqrt{n}}, \mu_0 + \frac{z_{\frac{\alpha}{2}} \cdot \sigma_0}{\sqrt{n}} \right), \text{ 拒绝域为 } (-\infty, \lambda_1) \cup (\lambda_2, +\infty), \text{ 故}$$

$$\alpha = P\{\text{第一类错误}\} = P\{\text{弃真错误}\} = P\{\text{拒绝 } H_0 \mid H_0 \text{ 正确}\}$$

$$= P\{\bar{X} < \lambda_1 \mid \mu = \mu_0\} + P\{\bar{X} > \lambda_2 \mid \mu = \mu_0\}$$

$$= P\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma_0} < \frac{\sqrt{n}(\lambda_1 - \mu_0)}{\sigma_0}\right\} + 1 - P\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma_0} \leq \frac{\sqrt{n}(\lambda_2 - \mu_0)}{\sigma_0}\right\}$$

$$= 1 + \Phi\left(\frac{\sqrt{n}(\lambda_1 - \mu_0)}{\sigma_0}\right) - \Phi\left(\frac{\sqrt{n}(\lambda_2 - \mu_0)}{\sigma_0}\right)$$

$$\beta = P\{\text{第二类错误}\} = P\{\text{取伪错误}\} = P\{\text{接受 } H_0 \mid H_0 \text{ 不正确, } H_1 \text{ 正确}\}$$

$$= P\{\lambda_1 < \bar{X} < \lambda_2 \mid \mu = \mu_1\}$$

$$= P\left\{\frac{\sqrt{n}(\lambda_1 - \mu_0)}{\sigma_0} < \frac{\sqrt{n}(\bar{X} - \mu_1)}{\sigma_0} < \frac{\sqrt{n}(\lambda_2 - \mu_0)}{\sigma_0}\right\} = \Phi\left(\frac{\sqrt{n}(\lambda_2 - \mu_1)}{\sigma_0}\right) - \Phi\left(\frac{\sqrt{n}(\lambda_1 - \mu_1)}{\sigma_0}\right)$$