概率统计讲义

对应教材：何书元《概率论与数理统计》

2025年秋季学期

目录

介绍

课程介绍

- 掌握概率论和数理统计的基本数学知识。
- 训练用概率论和数理统计方法对实际问题进行数学建模的能力。
- 学会解决常见的统计分析问题。
- 是应用型很强的学科。

参考书

- 教材：何书元：《概率论与数理统计》，高等教育出版社，2006.
- 陈家鼎、刘婉如、汪仁官：《概率统计讲义》（第三版）, 高等教育出版社，2004.
- 何书元《概率论》，北京大学出版社，2005.
- 李贤平，《概率论基础》第三版，高等教育出版社, 2010; 李贤平，陈子毅, 《概率论基础学习指导书》，高等教育出版社，2011
- Sheldon M. Ross,《概率论基础教程》（A First Course in Probability），人民邮电出版社，2006, 郑忠国、詹从赞翻译
- 陈家鼎，郑忠国，《概率与统计》，北京大学出版社，2007
- 程士宏，《测度论与概率论基础》，北京大学出版社，2004
- Sheldon M. Ross, 《应用随机过程–概率模型导论》（Introduction to Probability Models），人民邮电出版社，2011，龚光鲁翻译
- 陈家鼎、孙山泽、李东风、刘力平，《数理统计学讲义》，高等教育出版社，第三版，2015 年。
- Robert V. Hogg, Joseph W. McKean and Allen T. Craig, Introduction to Mathematical Statistics(7th ed.), 机械工业出版社，2012。

概率论的内容

- 随机事件与概率；
- 随机变量及其概率分布；
- 多维随机变量及其概率分布；
- 随机变量的数字特征；
- 大数定律及中心极限定理。

数理统计的内容

- 描述统计；
- 参数估计；
- 假设检验；
- 回归分析.

第一章　古典概型和概率空间

1.1　试验与事件

第一章介绍

- 在考虑一个 (未来) 事件是否会发生的时候, 人们常关心该事件发生的可能性的大小.
- 就像用尺子测量物体的长度、我们用概率测量一个未来事件发生的可能性大小.
- 将概率作用于被测事件就得到该事件发生的可能性大小的测量值.
- 为了介绍概率, 需要先介绍试验和事件.

试验

- 我们把按照一定的想法去作的事情称为随机试验. 随机试验的简称是**试验** (experiment).
- 下面都是试验的例子.
- 掷一个硬币, 观察是否正面朝上,
- 掷两枚骰子, 观察掷出的点数之和,
- 在一副扑克牌中随机抽取两张, 观察是否得到数字相同的一对,
- 有 7 个运动员参加 100 米短跑比赛, 观测比赛结果的名次排列,
- 乘电梯从一楼上到 9 楼, 观测电梯一共停了几次;
- 观测放学回家的路上所用的时间;
- 观测航天器发射的成功与否;

- 观察明天的最高气温;
- 考察某商场在一天内来到的顾客数量;
- 观测下次概率统计课有多少同学迟到.
- 观察 2003 年爆发的非典型肺炎案例首次下降到零的日期.
- 在概率论的语言中, 试验还是指对试验的一次观测或试验结果的测量过程.

样本空间

- 投掷一枚硬币, 用 ω_+ 表示硬币正面朝上, 用 ω_- 表示硬币反面朝上, 则试验有两个可能的结果：ω_+ 和 ω_-. 我们称 ω_+ 和 ω_- 是**样本点**, 称样本点的集合 $\Omega = \{\omega_+, \omega_-\}$ 为试验的**样本空间**.
- 投掷一枚骰子, 用 1 表示掷出点数 1, 用 2 表示掷出点数 2, $\cdots$, 用 6 表示掷出点数 6. 试验的可能结果是 1, 2, 3, 4, 5, 6. 我们称这 6 个数是试验的**样本点**. 称样本点的集合

  $$\Omega = \{\omega \,|\, \omega = 1, 2, \cdots, 6\}$$

  是试验的样本空间.
- 为了叙述的方便和明确, 下面把一个特定的试验称为试验 S.
- **样本点** (sample point): 称试验 S 的可能结果为样本点, 用 ω 表示.
- **样本空间** (sample space): 称试验 S 的样本点构成的集合为样本空间, 用 Ω 表示. 于是

  $$\Omega = \{\omega \mid \omega \text{ 是试验 } S \text{ 的样本点}\}.$$

事件

- 投掷一枚骰子的样本空间是

  $$\Omega = \{\omega \,|\, \omega = 1, 2, \cdots, 6\}.$$

- 用集合 $A = \{3\}$ 表示掷出 3 点, 则 A 是 Ω 的子集. 我们称 A 是**事件**.
- 掷出 3 点, 就称事件 A 发生, 否则称事件 A 不发生.

- 用集合 $B = \{2, 4, 6\}$ 表示掷出偶数点, B 是 Ω 的子集, 我们也称 B 是**事件**.
- 当掷出偶数点, 称事件 B 发生, 否则称事件 B 不发生. 事件 B 发生和掷出偶数点是等价的.
- **事件** (event): 设 Ω 是试验 S 的样本空间. 当 Ω 中只有有限个样本点时, 称 Ω 的子集为事件. 当试验的样本点 (试验结果) ω 落在 A 中, 称事件 A 发生, 否则称 A 不发生.
- 按照上述约定, 子集符号 $A \subset \Omega$ 表示 A 是事件. 通常用大写字母 A, B, C, D 或 $A_1, A_2, \cdots, B_1, B_2, \cdots$ 等表示事件.
- 用 $\overline{A} = \Omega - A$ 表示集合 A 的**余集**. 则事件 A 发生和样本点 $\omega \in A$ 是等价的, 事件 A 不发生和样本点 $\omega \in \overline{A}$ 是等价的.
- 空集 ϕ 是 Ω 的子集. 由于 ϕ 中没有样本点, 永远不会发生, 所以称 ϕ 是**不可能事件**. Ω 也是样本空间 Ω 的子集, 包含了所有的样本点, 因而总会发生. 我们称 Ω 是**必然事件**.

例 1.1: 投掷两枚硬币

- 投掷两枚硬币, 写出试验的样本点和样本空间.
- **解** 用 H(head) 表示硬币正面朝上, 用 T(tail) 表示硬币反面朝上,
- 试验一共有 4 个样本点, 他们是
  - HH
  - HT
  - TH
  - TT
- 样本空间是 $\Omega = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$.
- 注意, HT 和 TH 是不同的样本点.

例 1.2: 播音员选择

- **例 1.2** 某电视台要招聘播音员, 现在有三位符合条件的女士和两位符合条件的男士前来应聘.

  (1) 写出招聘男女播音员各一名的样本空间和样本点,

(2) 写出招聘两名播音员的样本空间 Ω 和事件 A=“招聘到两名女士”.

- **解** 本“试验”是招聘播音员. 用 W_1, W_2, W_3 分别表示第 1,2,3 位女士, 用 M_1,M_2 分别表示第 1,2 位男士.
- 用 W_1M_1 表示招聘到第 1 位女士和第 1 位男士, 用 W_1M_2 表示招聘到第 1 位女士和第 2 位男士, $\cdots$.
- (1) 招聘男女播音员各一名时, 样本空间是

$$\Omega = \{ W_1M_1,\ W_1M_2,\ W_2M_1,\ W_2M_2,\ W_3M_1,\ W_3M_2 \}.$$

  Ω 中的元素是样本点.

- (2) 招聘两名播音员时, 样本空间是

$$\Omega = \{\ W_1W_2,\ W_1W_3,\ W_2W_3,\ W_1M_1,\ W_1M_2,\ W_2M_1,\ W_2M_2,\ W_3M_1,\ W_3M_2,\ M_1M_2\ \}.$$

- 招聘到两名女士的事件 $A = \{\ W_1W_2,\ W_1W_3,\ W_2W_3\ \}$.

事件与集合

- 当 A, B 都是事件, 则

$$A \cup B,\quad A \cap B, A - B \triangleq A \cap \overline{B}$$

  都是事件. 也就是说事件经过集合运算得到的结果还是事件.(图示)

- 我们也用 AB 表示 $A \cap B$. 当 $AB = \phi$ 时, 也用 $A + B$ 表示 $A \cup B$.
- 当事件 $AB = \phi$, 称事件 A, B **不相容**. 特别称 $\overline{A}$ 为 A 的**对立事件** 或**逆事件**.
- 如果多个事件 $A_1, A_2, \ldots$ 两两不相容: $A_iA_j = \phi$, $i \neq j$, 就称他们**互不相容**.
- 注意，互不相容与后面要讲到的“独立”是完全不同的概念。
- 从以上的叙述看出, 从集合角度看，**样本空间 Ω 是由试验 S 的可能结果构成的全集, 样本点 ω 就是 Ω 的元素, 事件 A 就是 Ω 的子集.**
- 事件的运算符号和集合的运算符号也是相同的, 例如:

(1) $A = B$ 表示事件 A, B 相等,

(2) $A \cup B$ 发生 等价于 至少 A, B 之一发生,

(3) $A \cap B$ (或 AB) 发生 等价于 A 和 B 都发生,

(4) $\cup_{j=1}^{n} A_j$ 发生表示至少有一个 $A_j (1 \leq j \leq n)$ 发生, $\cup_{j=1}^{\infty} A_j$ 发生表示至少有一个 $A_j (j = 1, 2, \cdots)$ 发生,

(5) $\cap_{j=1}^{n} A_j$ 发生表示所有的 $A_j (1 \leq j \leq n)$ 都发生. $\cap_{j=1}^{\infty} A_j$ 发生表示所有的 $A_j (j = 1, 2, \cdots)$ 都发生.

事件的运算

事件的运算公式就是集合的运算公式, 例如[1]:

1. $A \cup B = B \cup A$, $A \cap B = B \cap A$,
2. $A \cup (B \cup C) = A \cup B \cup C$, $A \cap (B \cap C) = A \cap B \cap C$,
3. $A(B \cup C) = (AB) \cup (AC)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
4. $A \cup B = A + \overline{A}B$, $A = AB + A\overline{B}$,
5. 对偶公式: $\overline{A \cup B} = \bar{A} \cap \bar{B}$, $\overline{A \cap B} = \bar{A} \cup \bar{B}$, 进而有 $\overline{\bigcup_{j=1}^{\infty} A_j} = \bigcap_{j=1}^{\infty} \overline{A}_j$, $\overline{\bigcap_{j=1}^{\infty} A_j} = \bigcup_{j=1}^{\infty} \overline{A}_j$.

其中的公式 4 和 5 是值得牢记的.

1.2 古典概型与几何概型

1.2.1 古典概型

古典概率模型

- 设 Ω 是试验 S 的样本空间. 对于 Ω 的事件 A, 我们用 $P(A)$ 表示 A 发生的可能性的大小, 称 $P(A)$ 是事件 A 发生的**概率**, 简称为 A 的概率.
- 概率是介于 0 和 1 之间的数, 描述事件发生的可能性的大小.
- 按照以上原则, 如果事件 A, B 发生的可能性相同, 则有 $P(A) = P(B)$. 如果事件 A 发生的可能性比 B 发生的可能性大 2 倍, 则有 $P(A) = 2P(B)$.

[1] 图示讲解

- 用 $^{\#}A$, $^{\#}\Omega$ 分别表示事件 A 和样本空间 Ω 中样本点的个数.
- **定义** 2.1 设试验 S 的样本空间 Ω 是有限集合, $A \subset \Omega$. 如果 Ω 的每个样本点发生的可能性相同, 则称

$$P(A) = \frac{^{\#}A}{^{\#}\Omega} \tag{2.1}$$

  为试验 S 下 A 发生的**概率**, 简称为事件 A 的概率.
- 能够用定义 2.1 描述的模型称为**古典概率模型**, 简称为古典概型.

概率的性质

- 因为 $^{\#}A \geq 0$, 当 $AB = \phi$ 时，$^{\#}(A+B) =^{\#} A +^{\#} B$, 所以从定义 (2.1) 可以得到概率 P 的以下性质:
- (1) $P(A) \geq 0$,
- (2) $P(\Omega) = 1$,
- (3) 如果 A, B 不相容, 则 $P(A+B) = P(A) + P(B)$.
- 从以上的性质再得到
- (4) 如果 $A_1, A_2, \cdots, A_n$ 互不相容, 则

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n),$$

- (5) $P(\phi) = 0$, $P(A) + P(\overline{A}) = 1$, $P(A) = P(AB) + P(A\overline{B})$.
- 实际上, 我们由 (3) 得到 $P(\phi) + P(\Omega) = P(\Omega)$, 于是 $P(\phi) = 0$, 由 $A + \overline{A} = \Omega$ 和 (3) 得到 $P(A) + P(\overline{A}) = 1$, 由 $AB + A\overline{B} = A$ 和 (3) 得到 $P(A) = P(AB) + P(A\overline{B})$.

利用古典概型计算概率

- 列出样本空间所有样本点，一定注意这些样本点应该是可能性完全相同的；
- 计算样本空间样本点个数；
- 计算事件 A 样本点个数；
- 用公式 (2.1) 计算 $P(A)$。

例 2.1

- 以下的例子中任取、随机抽取都是指等可能的抽取。假设硬币、骰子等是均匀的。
- 掷一个均匀的硬币, 用 A 表示正面朝上.
- $$\begin{aligned} {}^{\#}\Omega =& 2 \\ {}^{\#}A =& 1 \\ P(A) =& 1/2 \end{aligned}$$

例 2.2

- 掷一个均匀的骰子, 用 A 表示掷出奇数, B 表示掷出 5.
- ${}^{\#}\Omega = 6,\ {}^{\#}A = 3, {}^{\#}B = 1$。
- $$P(A) = \frac{3}{6}, \quad P(B) = \frac{1}{6}.$$

古典概型中的常用计数—加法原理

- 如果一个问题的做法分为两类，第一类有 n 种方法，第二类有 m 种方法，这两类没有重叠而且仅有此两类，则问题的做法共有 $n+m$ 种。
- 多类的情况类似。
- **例** 选班长时，可以从 15 个男生中选一个，也可以从 10 个女生中选一个，那么一共有 $15+10=25$ 种选法。

古典概型中的常用计数—乘法原理

- 如果一个问题要两步完成，第一步有 n 种做法，第二步有 m 种做法，则问题有 nm 种做法。
- 多步的情况类似。
- **例** 要选一个男生班长和一个女生班长组成领导核心，男生 15 人，女生 10 人，则问题的做法有 $15 \times 10 = 150$ 种做法。

古典概型中常用计数—有重复的排列数

- 从 n 个不同元素中有放回地每次随机抽取一个，共抽取 m 次，有序地记录结果，共有 n^m 种等可能的不同结果。
- **例** 掷骰子 3 次，记录每次结果，结果一共有 $6 \times 6 \times 6 = 6^3$ 种。
- **例** 从 52 张扑克牌中随机有放回地抽取并记录 3 次，结果共有 52^3 种。

古典概型中常用计数—排列数

- 从 n 个不同元素中无放回地每次随机抽取一个，共抽取 m 次 $(m \leq n)$, 有序地记录结果，共有

$$A_n^m = n(n-1)\dots(n-m+1) = \frac{n!}{(n-m)!}$$

  种等可能的不同结果。
- A_n^m 在有的教材中记为 P_n^m。
- **例** 从 52 张扑克牌中随机无放回地抽取 3 张，记录每次结果，结果有 $52 \times 51 \times 50 = A_{52}^3$ 种。

古典概型中常用计数—组合数

- 从 n 个不同元素中无放回地每次抽取一个，共抽取 m 次 $(m \leq n)$, 不计次序地记录结果（只要元素相同，不管次序是否相同都算是相同结果），共有

$$C_n^m = \frac{n(n-1)\dots(n-m+1)}{m!} = \frac{n!}{m!(n-m)!}$$

  种等可能的不同结果。
- **例** 从一副扑克牌的 4 张 A 中随机无放回抽取 2 张组成一手牌，不计次序。有 $C_4^2 = 4 \times 3/2 = 6$ 种结果。分别为

♠♡ ♠♢ ♠♣ ♡♢ ♡♣ ♢♣

古典概型中常用计数—分组方式数

- 将 n 个不同元素分成有序号的 k 组，要求第 i 组恰好有 n_i 个元素 $(i = 1, 2, \ldots, k)$，分组结果中同组的元素不考虑次序。则这样分组的所有不同分法个数为
$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}.$$
- 当随机分组时，这些分法是等可能的。
- 随机分组的方法是 n 个元素随机排列 ($n!$ 种排法)，然后前 n_1 个不计次序地归入 $i = 1$ 组，后续 n_2 个不计次序地归入 $i = 2$ 组，以此类推。
- **例** 10 个学生分成 A, B, C 三个组，分别有 3、3、4 人，组内不计次序。
- 分组方式个数为
$$\frac{10!}{3!3!4!} \triangleq \binom{10}{3, 3, 4}$$

古典概型中常用计数—可重复分组数

- 从 n 个不同的球中有放回地每次抽取一个，共抽取 m 次，结果不计次序。
- 共有 C_{n+m-1}^{m} 种不同的组合。
- 参考：何书元《概率论》§1.2。
- 用 0 和 1 组成的序列表示一个结果。
- 用 $n-1$ 个 1 分隔出 n 个组，1 表示组边界。这 n 个组是结果排序后球号 $1, 2, \ldots, n$ 的组。
- 每组内有若干个 0 表示该组个数，如果出现 11 则该组没有球，把 m 个 0 分配到各个组中。
- 这样，用长度为 $n + m - 1$ 的 0-1 向量表示一个结果，结果个数为 C_{n+m-1}^{n-1}(从 $n + m - 1$ 个二进制位中选择 1 的位置，即边界的位置)。
- 可重复分组数在随机分组时一般不是等可能的。
- 例如，从红、白两个球中有放回地抽取 2 次，计数这 2 次红球、白球个数。

- 共有 (红 0，白 2), (红 1, 白 1), (红 2, 白 0) 三种结果, 即 $C_{2+2-1}^2 = 3$ 中结果。随机抽取时 (红 1, 白 1) 概率为 $\frac{1}{2}$, (红 0, 白 2) 和 (红 2, 白 0) 的概率都是 $\frac{1}{4}$。

例 2.3

- **例 2.3** 掷两个骰子, 用 A 表示点数之和为 7. 计算 $P(A)$.
- **解** 用 (i, j) 表示第一个骰子的点数是 i, 第二个骰子的点数是 j. 则

$$\Omega = \{(i,j) \,|\, i,j = 1,2,\cdots,6\},$$
$$A = \{(i,j) \,|\, i+j = 7, i,j = 1,2,\cdots,6\}.$$

  Ω 中的样本点具有等可能性. 由 $^{\#}\Omega = 6^2$, $^{\#}A = 6$ 知道 $P(A) = 6/36 = 1/6$.

- **注意**: 这个概率空间是有次序的，如果取无次序的概率空间 (比如 (1,2) 和 (2,1) 看成同一样本点) 则概率空间中的样本点不是等可能的。

例 2.4

- **例 2.4** 在 4 个白球, 6 个红球中任取 4 个, 求取到 2 个白球和 2 个红球的概率.
- **解** “任取”指无放回等可能随机抽取。用 A 表示取到 2 个白球和 2 个红球. 由 $^{\#}\Omega = C_{10}^4$, $^{\#}A = C_4^2 C_6^2$ 得到

$$P(A) = \frac{C_4^2 C_6^2}{C_{10}^4} = 0.4286.$$

例 2.5

- **例 2.5** 将 52 张扑克 (去掉两张王牌) 随机地分给 4 家, 求每家都是同花色的概率.
- **解** 认为 52 张牌被等可能地分为 4 组, 求每组 13 张牌同花色的概率. 这时 $^{\#}\Omega = 52!/(13!)^4$, $^{\#}A = 4!$, 故

$$P(A) = \frac{^{\#}A}{^{\#}\Omega} = \frac{4!(13!)^4}{52!} = 4.4739 \times 10^{-28}.$$

- 这样的小概率事件你和你周围的人是不会遇到的.

例 2.6

- **例 2.6** N 件产品中有 N_i 件 $i(1 \le i \le k)$ 等品, 从中任取 n 件. 求 n 件中恰有 n_i 件 i $(1 \le i \le k)$ 等品的概率.
- **解** 从题意知 $N_1 + N_2 + \cdots + N_k = N$, $n_1 + n_2 + \cdots + n_k = n$. 用 Ω 表示试验的样本空间, 用 A 表示取出的 n 件中恰有 n_i 件 i 等品,
- 则
$$^{\#}\Omega = C_N^n,\ ^{\#}A = C_{N_1}^{n_1} C_{N_2}^{n_2} \cdots C_{N_k}^{n_k}$$
- 于是
$$P(A) = \frac{C_{N_1}^{n_1} C_{N_2}^{n_2} \cdots C_{N_k}^{n_k}}{C_N^n}.$$

例 2.7(生日问题)

- **例 2.7** (生日问题) 求 n 个人中至少有两个人同生日的概率.
- **解** 认为每个人的生日等可能地出现在 365 天中的任一天, 则样本空间 Ω 的元素数为 $^{\#}\Omega = 365^n$.
- 用 $\overline{A}$ 表示 n 个人的生日各不相同, 则做为 Ω 的子集 $^{\#}\overline{A} = A_{365}^n$.
- 要求的概率
$$p_n = P(A) = 1 - P(\overline{A}) = 1 - A_{365}^n/365^n.$$
- 这里和以后规定对 $k > n$, $A_n^k = C_n^k = 0$. 可以计算出以下结果:

n	20	30	40	50	60	70	80
p_n	0.411	0.706	0.891	0.970	0.994	0.999	0.9999

- 图 1.2.1 是 p_n 和 n 的关系图. 横坐标是 n, 纵坐标是 p_n. 可以看出, 当 n 增加时, p_n 增加得很快.

例 2.8

- **例 2.8** 设样本空间 Ω 有 n 个样本点, 在古典概率模型下证明
- (1) 如果 $A_1, A_2, \cdots$, 是事件, 则 $\bigcup_{j=1}^{\infty} A_j$ 是事件,
- **证明:** $\bigcup_{j=1}^{\infty} A_j \subset \Omega$, 所以 (1) 成立.

- (2) 对于互不相容的事件 $A_1, A_2, \cdots$,

$$P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j).$$

- **证明** (2) 因为 $^{\#}\Omega = n$, 所以只有有限个 A_j 非空. 设前 m 个 A_j 可能非空, 其余是空集. 则

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{m} A_j.$$

对 $j > m$, $P(A_j) = 0$. 于是用性质 (4) 得到

$$P(\bigcup_{j=1}^{\infty} A_j) = P(\bigcup_{j=1}^{m} A_j) = \sum_{j=1}^{m} P(A_j) = \sum_{j=1}^{\infty} P(A_j).$$

1.2.2 几何概型

欧式空间中的体积

- 用 $\mathbb{R}^r$ 表示 r 维欧式空间

$$\mathbb{R}^r = \{(x_1, x_2, \ldots, x_r) | x_i \in (-\infty, \infty), i = 1, 2, \ldots, r\}.$$

- 对于 $\mathbb{R}^r$ 的子集 A，用

$$m(A) = \int_A dx_1 dx_2 \ldots dx_r$$

表示 A 的体积。

- （更一般地，$m(A)$ 表示可测集 A 的测度，参见《实变函数》）

几何概率

- 用 Ω 表示试验 S 的样本空间，当 $\Omega \subset \mathbb{R}^r$ 时，称 Ω 的子集是**事件**。

- **定义** 设样本空间 Ω 的体积 $m(\Omega)$ 是正数，样本点等可能地落在 Ω 中（指 Ω 的体积相同的长方体事件发生的可能性相同），对于 $A \subset \Omega$，称

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

为事件 A 发生的概率，简称为 A 的概率。

- 这样的定义也满足 §1.2 中的非负性、全空间概率等于 1、可加性三个性质，从而性质 (4) 和 (5) 也成立。

例 (同心圆)

- 两个同心圆，大圆圆面为 Ω：半径 1m; 内部小圆圆面为 A：半径 0.5m。
- 落入 A 概率

$$P(A) = \frac{m(A)}{m(\Omega)} = \frac{\pi(0.5)^2}{\pi 1^2} = 0.25$$

- 落入 A 外的大圆的概率

$$P(\bar{A}) = 1 - P(A) = 0.75$$

例 (会面概率)

- 两人 1:00—2:00 间独立地随机到达某地会面，先到者仅等待 20 分钟。求会面概率。
- 用 x, y 表示两人分别的到达时间，则

$$\Omega = \{(x, y) : 0 \le x, y \le 60\},$$

  样本点 (x, y) 等可能地落在空间 Ω 内。
- A 表示两人相遇，则

$$A = \{(x, y) |\ |x - y| \le 20, (x, y) \in \Omega\}$$

- $m(\Omega) = 60^2$.
- $m(A)$ 用图示，A 的两条斜边为 $y = x \pm 20$，面积等于 60^2 减去两个三角形面积即 $2 \times \frac{1}{2} \times 40^2$，所以

$$m(A) = 60^2 - 40^2$$

- 概率

$$P(A) = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}.$$

1.3 概率的公理化和加法公式

1.3.1 概率的公理化

概率空间

- 古典概型只对样本空间 Ω 含有限个样本点, 且每个样本点发生的可能性相同的情况定义了概率. 下面将概率的定义进行推广.
- 设 Ω 是试验 S 的样本空间, 在实际问题中往往并不需要关心 Ω 的所有子集, 只要把关心的子集称为事件就够了. 但是事件必须是 Ω 的子集, 并且满足以下三个条件:
  - (a) Ω 是事件,
  - (b) A, B 是事件, 则 $A \cup B, A \cap B, A - B, \overline{A}$ 都是事件,
  - (c) 当 A_j 是事件, 则 $\bigcup_{j=1}^{\infty} A_j$ 是事件.
- 以后总假设上面的条件 (a), (b), (c) 成立. 由 (b) 知道有限个事件经过有限次运算后得到的结果仍然是事件.
- 满足条件 (a), (b), (c) 的事件的集合 $\mathscr{F}$ 叫做 σ 域或 σ 代数。
- 对于试验 S 的事件 A, 我们用 0,1 之间的数 $P(A)$ 表示事件 A 发生的可能性的大小. 对于每个事件 $A \in \mathscr{F}$, $P(A)$ 是一个实数. $P(A)$ 是事件 A 的函数.

概率及公理化

- **定义 3.1** 如果事件的函数 P 满足条件
  - (a) 非负性: 对于任何事件 A, $P(A) \geq 0$,
  - (b) 完全性: $P(\Omega) = 1$,
  - (c) 可列可加性: 对于互不相容的事件 $A_1, A_2, \ldots$, 有

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

  就称 P 是试验 S 的概率, 简称为概率, 称 $P(A)$ 是 A 的**概率** (probability).
- 我们称定义 3.1 中的 (a), (b), (c) 为概率的**公理化条件**.
- 不满足公理化条件的 P 不是概率.
- 条件 (c) 中的“可列”, 指集合的个数或运算的次数可以依次排列起来. 从例 2.8 知道, 古典概率模型中的 P 是概率.

概率的性质

- 概率的公理化条件并不直接告诉我们在实际问题中如何计算 $P(A)$. $P(A)$ 的计算要根据问题的条件和背景得到.
- 设 P 是试验 S 的概率, 则有以下的结果.
  - (1) 不可能事件的概率是 0: $P(\phi) = 0$,
  - (2) 有限可加性: 如果 $A_1, A_2, \cdots, A_n$ 是互不相容, 则

$$P(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P(A_j),$$

  - (3) 单调性: $B \subset A$, 则 $P(A) - P(B) = P(A - B) \geq 0$.

  (先证明有限可加性，$P(\emptyset) = 0$ 与单调性用有限可加性证明)

1.3.2 概率的加法公式

概率的加法公式

- 概率的有限可加性和可列可加性是概率 P 的最基本性质, 由此推出概率的加法公式.
- (4) $P(A \cup B) = P(A) + P(B) - P(AB)$,
- (5) 如果 $B \subset A$, 则 $P(A - B) = P(A) - P(B)$, $P(A) \geq P(B)$,
- (6) Jordan 公式: 设 $A_1, A_2, \cdots, A_n$ 是事件, 记

$$p_k = \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} P(A_{j_1} A_{j_2} \cdots A_{j_k})$$

  时, 有

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k-1} p_k. \tag{3.2}$$

例 3.1

- **例 3.1** 全班有 26 个人会打网球, 有 28 个人会打羽毛球, 他们中有 20 个人同时会打网球和羽毛球. 从全班的 40 名同学中任选一名, 计算他会打网球或会打羽毛球的概率.

- **解** 对任选出的同学, 用 A 表示他会打网球, 用 B 表示他会打羽毛球, 则 $A\cup B$ 表示他会打网球或会打羽毛球. 利用

$$P(A)=26/40,\ P(B)=28/40,\ P(AB)=20/40$$

得到

$$P(A\cup B)=P(A)+P(B)-P(AB)=\frac{26+28-20}{40}=0.85.$$

1.3.3 概率的连续性

概率的连续性

- 如果 $A_1\subset A_2\subset\cdots$, 就称事件序列 $\{A_j\}\equiv\{A_j\mid j=1,2,\cdots\}$ 是**单调增**的.
- 如果 $A_1\supset A_2\supset\cdots$, 就称事件序列 $\{A_j\}$ 是**单调减**的.
- 我们把单调增序列和单调减序列统称为单调序列.
- **定理 3.1** 设 $\{A_j\}$ 和 $\{B_j\}$ 是事件列.
  - (1) 如果 $\{A_j\}$ 是单调增序列, 则

$$P(\bigcup\nolimits_{j=1}^{\infty}A_j)=\lim_{n\to\infty}P(A_n).$$

  - (2) 如果 $\{B_j\}$ 是单调减序列, 则

$$P(\bigcap\nolimits_{j=1}^{\infty}B_j)=\lim_{n\to\infty}P(B_n).$$

- 通常称 $\bigcup_{j=1}^{\infty}A_j$ 为单调增序列 $\{A_j\}$ 的极限, 称 $\bigcap_{j=1}^{\infty}B_j$ 为单调减序列 $\{B_j\}$ 的极限.
- 定理 3.1 说明, A_j 的概率收敛到它的极限 $\bigcup_{j=1}^{\infty}A_j$ 的概率, B_j 的概率收敛到它的极限 $\bigcap_{j=1}^{\infty}B_j$ 的概率. 所以称概率具有**连续性**.

1.4 条件概率和乘法公式

例 4.1：掷骰子的条件概率

- **例 4.1** 掷一个骰子, 已知掷出了偶数点, 求掷出的是 2 的概率.
- 用 A 表示掷出偶数点, B 表示掷出 2.

- 已知 A 发生后试验的条件已经改变. 在新的试验条件下 A 成为样本空间, A 的样本点具有等可能性, B 是 A 的子集, $^{\#}A = 3$, $^{\#}B = 1$. 所以, 用 $P(B|A)$ 表示要求的概率时,

$$P(B|A) = \frac{^{\#}B}{^{\#}A} = \frac{1}{3}.$$

- 我们称 $P(B|A)$ 是已知 A 发生的条件下, B 发生的概率.

例 4.2: 扑克牌的条件概率

- **例 4.2** 在 52 张扑克中任取一张, 已知抽到草花的条件下, 求抽到的是草花 5 的概率.

- **解** 设 A=“抽到草花”, B=“抽到草花 5”. 按例 4.1 的方法有 $P(B|A) = {}^{\#}B/{}^{\#}A = 1/13$.

条件概率

- 设 A, B 是事件, 以后总用 $P(B|A)$ 表示已知 A 发生的条件下, B 发生的条件概率, 简称为**条件概率** (conditional probability).

- 下面是条件概率的计算公式.

- **条件概率公式**: 如果 $P(A) > 0$, 则

$$P(B|A) = \frac{P(AB)}{P(A)}. \tag{4.1}$$

- 可以对古典概型给出 (4.1) 的证明: 设试验 S 的样本空间是 Ω, A, B 是事件, $P(A) > 0$. 已知 A 发生后试验的条件已经改变. 在新的试验条件下 A 成为样本空间, A 的样本点具有等可能性. 已知 A 发生后, AB 是 A 的子集. 利用古典概型的定义知道

$$P(B|A) = P(AB|A) = \frac{^{\#}(AB)}{^{\#}A} = \frac{^{\#}(AB)/^{\#}\Omega}{^{\#}A/^{\#}\Omega} = \frac{P(AB)}{P(A)}.$$

例 4.3: 扑克牌问题

- 在计算条件概率时, 公式 (4.1) 有时会带来许多的方便. 但有时根据问题的特点可以直接得到结果.

- **例 4.3** 将一副扑克牌的 52 张随机均分给四家, 设 A= 东家得到 6 张草花, B= 西家得到 3 张草花, 求 $P(B|A)$.

- **解** 四家各有 13 张牌, 可以认为东家先取 13 张后西家再取 13 张。
- 在 A 发生的条件下, 西家的总可能取法为 C_{39}^{13}，B 发生要求西家取法为 $C_7^3C_{32}^{10}$, 用古典概型

$$P(B|A)=\frac{C_7^3C_{32}^{10}}{C_{39}^{13}}$$

.

例 4.4: 条件概率是公理化概率

- **例 4.4** 设 $P(A)>0$, 对于任何事件 B, 定义 $P_A(B)=P(B|A)$. 则
- (1) P_A 是概率,
- (2) 对于事件 B, C, 当 $P(AB)>0$ 时,

$$P_A(C|B)=P(C|AB). \tag{4.2}$$

- (证明略)

乘法公式

- **乘法公式:** 设 $A,B,A_1,A_2,\ldots,A_n$ 是事件, 则
- (1) $P(AB)=P(A)P(B|A)$,
- (2) 当 $P(A_1A_2\ldots A_{n-1})\neq 0$, 有

$$P(A_1A_2\ldots A_n)=P(A_1)P(A_2|A_1)\ldots P(A_n|A_1A_2\ldots A_{n-1}). \tag{4.3}$$

- **证明** 将条件概率公式 (4.1) 用于等式右边的条件概率就得到证明. (手写)

例 4.5: 官员受贿问题

- **例 4.5** (官员受贿问题) 某官员第 1 次受贿没被查处的概率是 $q_1=98/100=0.98$. 第 1 次没被查处后, 第 2 次受贿没被查处的概率是 $q_2=96/98=0.9796$, $\cdots$. 前 $j-1$ 次没被查处后, 第 j 次受贿不被查处的概率是 $q_j=(100-2j)/(100-2(j-1))$, $\cdots$. 求他受贿 n 次还不被查处的概率 p_n.
- **解** 用 A_j 表示该官员第 j 次受贿没被查处, 则 $A_1A_2\cdots A_n$ 表示受贿 n 次还不被查处.

- $$\begin{aligned} p_n =& P(A_1A_2\dots A_n) \\ =& P(A_1)P(A_2|A_1)\dots P(A_n|A_1A_2\dots A_{n-1}) \\ =& q_1q_2\cdots q_n \\ =& \frac{98}{100}\frac{96}{98}\cdots\frac{100-2(n-1)}{100-2(n-2)}\frac{100-2n}{100-2(n-1)} \\ =& \frac{100-2n}{100} \\ =& 1-\frac{n}{50}. \end{aligned}$$

- 容易计算

$$p_{10}=0.8,\quad p_{20}=0.6,\quad p_{30}=0.4\quad p_{40}=0.2,\quad p_{50}=0.$$

- (如果假设 $q_1=q_2=\cdots=0.98$, 有 $p_n=0.98^n$, 于是得到 $p_{10}=0.817$, $p_{20}=0.668$, $p_{30}=0.545$, $p_{40}=0.446$, $p_{50}=0.364$, $p_{60}=0.298$.)

1.5 事件的独立性

两个事件独立

- 设 A 是试验 S_1 下的事件, B 是试验 S_2 下的事件, 且 A 的发生与否不影响 B 的发生. 用公式表述出来就是 $P(B|A)=P(B)$. (设 $P(A)>0$)
- 再用乘法公式得到 $P(AB)=P(A)P(B|A)=P(A)P(B)$. 此式表示事件 A, B 相互独立, 不要求 $P(A)>0$.
- **定义 5.1** 如果事件 A, B 满足 $P(AB)=P(A)P(B)$, 就称 A, B **相互独立**, 简称为 A, B **独立** (independent).
- 不可能事件, 必然事件与任何事件独立. 这是因为 $P(\phi A)=P(\phi)P(A)=0$, $P(\Omega A)=P(\Omega)P(A)=P(A)$ 总成立.
- 当 $0<P(A)<1$ 时, A, B 独立当且仅当 $P(B|A)=P(B|\bar{A})=P(B)$, 即 B 的概率不受已知 A 是否发生的影响。

例 5.1: 独立的试验

- **例 5.1** 用 Ω_1 表示试验 S_1 的样本空间, 用 Ω_2 表示 S_2 的样本空间.
- 如果试验 S_1 和 S_2 是独立进行的, 可以证明试验 S_1 的事件和试验 S_2 的事件是相互独立的.

例 5.2: 长方形等分

- **例 5.2** 两线段将长方形 Ω 四等分, 得到 E_1, E_2, E_3, E_4.
-

E_1	E_2
E_3	E_4

- 设 $A = \mathrm{E}_1 \cup \mathrm{E}_2$, $B = \mathrm{E}_1 \cup \mathrm{E}_3$, $C = \mathrm{E}_1 \cup \mathrm{E}_4$.
- 在 Ω 中任取一点, 则

$$P(AB) = P(A)P(B) = 1/4,$$
$$P(AC) = P(A)P(C) = 1/4,$$
$$P(BC) = P(B)P(C) = 1/4.$$

  于是 A, B, C 两两独立.

定理 5.1

- **定理 5.1** A, B 独立当且仅当 $\overline{A}$, B 独立.
- **证明** 只需由 A, B 独立证明 $\overline{A}$, B 独立. 当 A, B 独立, 有

$$P(\overline{A}B) = P(B) - P(AB) = P(B) - P(A)P(B) = P(\overline{A})P(B).$$

  于是 $\overline{A}$, B 独立.

多个事件的相互独立

- **定义 5.2** (1) 称事件 A_1, A_2, $\cdots$, A_n **相互独立**, 如果对任何 $1 \le j_1 < j_2 < \cdots < j_k \le n$,

$$P(A_{j_1}A_{j_2}\cdots A_{j_k}) = P(A_{j_1})P(A_{j_2})\ldots P(A_{j_k})$$

- (2) 称事件 A_1, A_2, $\cdots$ **相互独立**, 如果对任何 $n \ge 2$, 事件 A_1, A_2, $\cdots$, A_n 相互独立.
- (3) 称 $\{A_n\}$ 是**独立事件列**, 如果 A_1, A_2, $\cdots$ 相互独立.

例 5.3: 三个事件相互独立

- **例 5.3** 事件 A,B, C 相互独立当且仅当他们两两独立, 并且 $P(ABC) = P(A)P(B)P(C)$.

例 5.4: 性质

例 5.4 设 $A_1, A_2, \cdots, A_n$ 相互独立, 则有如下的结果.

- (1) 对 $1 \le j_1 < j_2 < \cdots < j_k \le n$, $A_{j_1}, A_{j_2}, \cdots, A_{j_k}$ 相互独立,
- (2) 用 B_i 表示 A_i 或 $\overline{A}_i$, 则 $B_1, B_2, \cdots, B_n$ 相互独立,
- (3) $(A_1A_2), A_3, \cdots, A_n$ 相互独立;
- (4) $(A_1 \cup A_2), A_3, \cdots, A_n$ 相互独立.

事实上，把 $A_1, A_2, \ldots, A_n$ 分为 k 个组，每个组内的事件作并、交、差运算后得到的事件 $B_1, B_2, \ldots, B_k$ 仍是相互独立的。

例 5.5

- **例 5.5** 例 5.2 中的 A,B,C 两两独立但非相互独立。
- 因为

$$P(ABC) = 1/4, \quad P(A)P(B)P(C) = 1/8.$$

例 5.6: 高炮

- **例 5.6** 每门高炮击中飞机的概率是 0.3, 要以 99% 的把握击中飞机, 需要几门高炮.
- **解** 用 A_i 表示第 i 门高炮击中目标. 设需要 n 门高炮, 则要求 n 满足

$$P(\bigcup_{j=1}^{n} A_j) = 1 - P(\bigcap_{j=1}^{n} \overline{A}_j) = 1 - (0.7)^n \ge 0.99.$$

- 由 $n \ln 0.7 \le \ln(1 - 0.99)$ 解出 $n \ge 12.9114$. 于是取 $n = 13$.

例 5.7: 明青花

- **例 5.7** 明青花 (瓷) 享有盛誉. 设一只青花盘在一年中被失手打破的概率是 0.03.
  - (1) 计算一只弘治 (1488-1505) 时期的青花麒麟 (图案) 盘保留到现在 (约 500 年) 的概率,
  - (2) 如果弘治年间生产了 1 万件青花麒麟盘, 计算这 1 万件至今都已经被失手打破的概率.

- **解**

  (1) 用 A_i 表示该盘在第 i 年没被打破, 则至今没被打破的概率是

$$\begin{aligned} p =& P(A_1A_2\cdots A_{500}) \\ =& P(A_1)P(A_2|A_1)\cdots P(A_{500}|A_1A_2\cdots A_{499}) \\ =& (1-0.03)^{500} \\ =& 2.43\times 10^{-7}, \end{aligned}$$

  被失手打破的概率是 $q = 1 - p = 0.999999756$.

- (2) 用 B_j 表示第 j 件至今已被打破, $m = 10000$, 则 $B_1, B_2, \cdots, B_m$ 相互独立.

- 这 1 万件至今都已经被失手打破的概率是

$$q_1 = P(\bigcap\nolimits_{j=1}^{m} B_j) = \prod_{j=1}^{m} P(B_j) = q^m = 0.99757.$$

- 有这类青花麒麟盘流传至今的概率是 $p_1 = 1 - q_1 = 0.00243$.

- 如果当时生产了五十万件, 则有这类青花麒麟盘流传至今的概率是 $p_{50} = 0.1149$.

- 如果当时生产了五百万件, 则有这类青花麒麟盘流传至今的概率是 $p_{500} = 0.7048$.

- 当然，这个模型里每年失手打破的概率都是 0.03 的假设过于粗糙，实际上随着现存总数的减少保护必然加强，打破概率变得很小。

1.6 全概率公式与 Bayes 公式

1.6.1 全概率公式

全概率公式

- **定理 6.1**(全概率公式) 如果事件 $A_1, A_2, \cdots, A_n$ 互不相容, $B \subset \bigcup_{j=1}^{n} A_j$, 则

$$P(B) = \sum_{j=1}^{n} P(A_j)P(B|A_j). \tag{6.1}$$

- **证明:** 因为 $B = B(\bigcup_{j=1}^{n} A_j) = \bigcup_{j=1}^{n}(BA_j)$, 用概率的有限可加性和乘法公式得到

$$\begin{aligned} P(B) =& P(B(\bigcup\nolimits_{j=1}^{n} A_j)) \\ =& P(\bigcup\nolimits_{j=1}^{n}(BA_j)) \\ =& \sum_{j=1}^{n} P(BA_j) \\ =& \sum_{j=1}^{n} P(A_j)P(B|A_j). \end{aligned}$$

全概率公式—完备事件组

- 全概率公式容易推广到可列个事件的情况 (见习题 1.14)).
- 如果事件 $A_1, A_2, \cdots, A_n$ 互不相容, $\bigcup_{j=1}^{n} A_j = \Omega$, 则称 $A_1, A_2, \cdots, A_n$ 是**完备事件组**, 这时 (6.1) 对任何事件 B 成立.
- A 和 $\overline{A}$ 总构成完备事件组, 所以总有

$$P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A}). \tag{6.2}$$

例 6.1(抽签问题)

- **例 6.1** (抽签问题) n 个签中有 m 个标有“中”, 证明无放回依次随机抽签时, 第 j 次抽中的概率是 m/n.
- **解** 用归纳法. 用 A_j 表示第 j 次抽中, 则对一切 m, n, 当 $m \le n$ 时, 有 $P(A_1) = m/n$.
- 设对一切 m, n, 当 $m \le n$ 时, 有 $P(A_{j-1}) = m/n$, 则有

$$P(A_j|A_1) = \frac{m-1}{n-1}, \quad P(A_j|\overline{A}_1) = \frac{m}{n-1}.$$

- 于是有

$$\begin{aligned} P(A_j) =& P(A_1)P(A_j|A_1) + P(\overline{A}_1)P(A_j|\overline{A}_1) \\ =& \frac{m}{n}\,\frac{m-1}{n-1} + \frac{n-m}{n}\,\frac{m}{n-1} \\ =& \frac{m}{n}, \quad 1 \le j \le n. \end{aligned}$$

例 6.2(敏感问题调查)

- **例 6.2** (敏感问题调查) 在调查家庭暴力 (或婚外恋、服用兴奋剂、吸毒等敏感问题) 所占家庭的比例 p 时, 被调查者往往不愿回答真相, 这使得调查数据失真.

- 为得到实际的 p 同时又不侵犯个人隐私, 调查人员将袋中放入比例是 p_0 的红球和比例是 $q_0 = 1 - p_0$ 的白球.

- 被调查者在袋中任取一球窥视后放回, 并承诺取得红球就讲真话, 取到白球就讲假话.

- 被调查者只需在匿名调查表中选“是”(有家庭暴力) 或“否”, 然后将表放入投票箱.

- 没人能知道被调查者是否讲真话和回答的是什么. 如果声称有家庭暴力的家庭比例是 p_1, 求真正有家庭暴力的比例 p.

- **解** 对任选的一个家庭, 用 B 表示回答“是”, 用 A 表示实际“是”. 利用全概率公式得到

$$\begin{aligned} p_1 =& P(B) \quad (\text{回答 “是”}) \\ =& P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) \\ =& p_0 P(A) + q_0(1 - P(A)) \\ & \big(P(B|A) \text{ 即讲真话概率}, P(B|\overline{A}) \text{ 等于讲假话概率}\big) \\ =& pp_0 + q_0(1-p) = q_0 + (p_0 - q_0)p. \end{aligned}$$

- 于是只要 $p_0 \neq q_0$, 则

$$p = P(A) = \frac{p_1 - q_0}{p_0 - q_0}.$$

- 实际问题中, p_1 是未知的, 需要经过调查得到. 假定调查了 n 个家庭, 其中有 k 个家庭回答“是”, 则可以用 $\hat{p}_1 = k/n$ 估计 p_1, 于是可以用

$$\hat{p} = \frac{\hat{p}_1 - q_0}{p_0 - q_0}$$

估计 p.

- 如果袋中装有 30 个红球，50 个白球，调查了 320 个家庭，其中有 195 个家庭回答“是”，则

$$\begin{aligned} &p_0 = 3/8, \quad q_0 = 5/8, \\ &\hat{p}_1 = 195/320, \\ &\hat{p} = \frac{195/320 - 5/8}{3/8 - 5/8} = 6.25\%. \end{aligned}$$

- 可以证明 $|p_0 - q_0|$ 越大, 得到的结论越可靠. 但是 $|p_0 - q_0|$ 太大时, 调查方案不易让被调查者接受.

例 6.3(赌徒破产模型)

- **例 6.3** (赌徒破产模型)　甲有本金 a 元, 决心再赢 b 元停止赌博. 设甲每局赢的概率是 $p = 1/2$, 每局输赢都是一元钱, 甲输光后停止赌博, 求甲输光的概率 $q(a)$.

- **解**　用 A 表示甲第一局赢, 用 B_k 表示甲有本金 k 元时最后输光.

- 由题意, $q(0) = 1$, $q(a+b) = 0$, 并且

$$\begin{aligned} q(k) =& P(B_k) \\ =& P(A)P(B_k|A) + P(\overline{A})P(B_k|\overline{A}) \\ =& \frac{1}{2}P(B_{k+1}) + \frac{1}{2}P(B_{k-1}) \\ =& \frac{1}{2}q(k+1) + \frac{1}{2}q(k-1). \end{aligned}$$

- 于是有 $2q(k) = q(k+1) + q(k-1)$.

- 从而得到

$$q(k+1) - q(k) = q(k) - q(k-1) = \cdots = q(1) - q(0) = q(1) - 1.$$

- 上式两边对 $k = n-1, n-2, \cdots, 0$ 求和后得到,

$$q(n) - 1 = n(q(1) - 1). \tag{6.3}$$

- 取 $n = a + b$, 得到

$$0 - 1 = (a+b)(q(1) - 1), \quad q(1) - 1 = -1/(a+b).$$

- 最后由 (6.3) 得到:

$$q(a) = 1 + a(q(1) - 1) = 1 - \frac{a}{a+b} = \frac{b}{b+a}. \tag{6.4}$$

- (6.4) 说明, 当甲的本金 a 有限, 则贪心 b 越大, 输光的概率越大, 如果一直赌下去 ($b \to \infty$), 必定输光.

1.6.2 Bayes 公式

Bayes 公式

- **定理 6.2**(Bayes 公式) 如果事件 $A_1, A_2, \cdots, A_n$ 互不相容, $B \subset \bigcup_{j=1}^{n} A_j$, 则 $P(B) > 0$ 时, 有

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}, \quad 1 \le j \le n. \tag{6.5}$$

- **证明** 由条件概率公式和全概率公式得到

$$P(A_j|B) = \frac{P(A_jB)}{P(B)} = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}, \quad 1 \le j \le n.$$

- 值得指出的是, 分子总是分母中的一项.

- 当 $A_1, A_2, \cdots, A_n$ 是完备事件组, $P(B) > 0$ 时, (6.5) 成立.

- Bayes 公式也可以推广到可列个事件的情况 (见习题 1.21).

- 最常用到的 Bayes 公式是当 $P(B) > 0$,

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})}. \tag{6.6}$$

例 6.4(疾病普查问题)

- **例 6.4** (疾病普查问题) 一种新方法对某种特定疾病的诊断准确率是 90%(有病被正确诊断和没病被正确诊断的概率都是 90%). 如果群体中这种病的发病率是 0.1% , 甲在身体普查中被诊断患病, 问甲的确患病的概率是多少?

- **解** 设 A= 甲患病, B = 甲被诊断有病.

- 根据题意, $P(A) = 0.001$,

$$P(B|A) = 0.9, \quad P(B|\overline{A}) = 0.1,$$

- 于是, 用公式 (6.6) 得到

$$\begin{aligned} P(A|B) =& \frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})} \\ =& \frac{0.001\times 0.9}{0.001\times 0.9+0.999\times 0.1} \\ =& \frac{9}{9+999}=0.0089<1\%. \end{aligned}$$

- 没有病的概率 $P(\overline{A}|B)=0.9911>99\%$.
- 造成这个结果的原因是发病率较低和诊断的准确性不够高.
- 如果甲复查时又被诊断有病, 则他的确有病的概率将增加到 7.5%.
- 如果人群的发病率不变, 诊断的准确率提高到 99%, 可以计算出 $P(A|B)=9.02\%$.

例 6.5(吸烟与肺癌问题)

- **例 6.5** (吸烟与肺癌问题) 1950 年某地区曾对 50-60 岁的男性公民进行调查. 肺癌病人中吸烟的比例是 99.7%, 无肺癌人中吸烟的比例是 95.8%. 如果整个人群的发病率是 $p=10^{-4}$, 求吸烟人群中的肺癌发病率和不吸烟人群中的肺癌发病率.
- **解** 引入 $A=$ 有肺癌, $B=$ 吸烟, 则

$$\begin{aligned} P(A) =& 10^{-4}, \\ P(B|A) =& 99.7\%, \\ P(B|\overline{A}) =& 95.8\%. \end{aligned}$$

- 利用公式 (6.6) 得到:

$$\begin{aligned} P(A|B) =& \frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})} \\ =& \frac{10^{-4}\times 99.7\%}{10^{-4}\times 99.7\%+(1-10^{-4})\times 95.8\%} \\ =& 1.0407\times 10^{-4}. \\ P(A|\overline{B}) =& \frac{P(A)P(\overline{B}|A)}{P(A)P(\overline{B}|A)+P(\overline{A})P(\overline{B}|\overline{A})} \\ =& \frac{10^{-4}\times(1-99.7\%)}{10^{-4}\times(1-99.7\%)+(1-10^{-4})\times(1-95.8\%)} \\ =& 7.1438\times 10^{-6}. \end{aligned}$$

- 于是,
$$\frac{\text{吸烟人群的发病率}}{\text{不吸烟人群的发病率}} = \frac{P(A|B)}{P(A|\overline{B})} = 14.57.$$
- 结论: 吸烟人群的肺癌发病率是不吸烟人群的肺癌发病率的 14.57 倍.

例 6.6(肇事车判定)

- **例 6.6** 某城市夏利牌出租车占 85%, 富康牌出租车占 15%. 这两种出租车都是红色, 富康出租车略大一些, 每辆车肇事的概率相同.
- 在一次出租车的交通肇事逃逸案件中, 有证人指证富康车肇事. 为了确定是否富康车肇事, 在肇事地点和相似的能见度下警方对证人辨别出租车的能力进行了测验, 发现证人正确识别富康车的概率是 90%, 正确识别夏利车的概率是 80%.
- 如果证人没有撒谎, 求富康车肇事的概率.
- **解:** 用 A 表示证人看见富康车肇事, 用 B 表示富康车肇事, 则 $\overline{B}$ 表示夏利车肇事, 并且
- $$P(B) = 0.15,\ P(A|B) = 0.9,\ P(A|\overline{B}) = 1 - 0.8.$$
- 要求的概率是 $P(B|A)$. 用 Bayes 公式得到
$$\begin{aligned} P(B|A) =& \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})} \\ =& \frac{0.9 \times 0.15}{0.9 \times 0.15 + (1 - 0.8) \times 0.85} \\ =& 44.26\%. \end{aligned}$$
- 这个概率看起来很小, 但是在没有证人的情况下, 富康车肇事的概率更小, 是 15%.

1.7 概率与频率

概率与频率

- 古典概型只对等可能的情况定义了概率, 为了能够描述更复杂的试验, 很多学者使用概率的频率定义.

- 设 A 是试验 S 的事件. 在相同的条件下将试验 S 独立地重复 N 次, 我们称
$$f_N = \frac{N\text{次试验中 } A \text{ 发生的次数}}{N}$$
是 N 次独立重复试验中, 事件 A 发生的**频率** (frequency).

- 理论和试验都证明, 当 $N \to \infty$, f_N 会收敛到一个数 $P(A)$. 我们称 $P(A)$ 为事件 A 在试验 S 下发生的概率, 简称为 A 的概率.

- (Flash 演示)

第二章　随机变量和概率分布

2.1　随机变量

随机变量——引入

- 事件是用来描述随机试验的某些现象是否出现的, 要说明比较复杂的试验结果, 就需要定义许多事件.
- 为了更深入地研究随机现象, 就要建立数学模型，**随机变量**是随机现象的最基本的数学模型，我们用随机变量的值表示随机试验的结果。
- 如果用 X 表示明天的最高气温, $\{X \le 30\}$ 就表示明天的最高气温不超过 30°C, 由于 X 的取值在今天无法确定, 所以称 X 是**随机变量** (random variable).

例 1.1: 骰子点数

- **例 1.1**　掷一个骰子, 样本空间是

$$\Omega = \{\, \omega \mid \omega = 1, 2, \cdots, 6 \,\}.$$

- 用 X 表示掷出的点数, 称 X 是随机变量.
- $\{X \le 3\}$ 表示掷出的点数不超过 3, 并且

$$\{X \le 3\} = \{\, \omega \mid \omega = 1, 2, 3\}$$

是事件。

- 将 X 视为 Ω 上的函数,

$$X(\omega) = \omega, \quad \omega \in \Omega,$$

则

$$\{X \le j\} = \{\omega \mid \omega = 1, 2, \cdots, j\}.$$

是事件。

例 1.2: 扑克牌点数

- **例 1.2** 在一副扑克的 52 张中任取一张, 样本空间的每个点表示一张扑克.
- 用 X 表示所取扑克的大小, 则 $X = 3$ 表示所取到的扑克是 3.
- 将 X 视为样本空间上的函数, 则

  $$\{X = 3\} \equiv \{\ \omega \mid X(\omega) = 3\ \} = \{\ \text{草花 3, 黑桃 3, 红桃 3, 方块 3}\ \}.$$

  是事件。
- 我们称 X 是随机变量.

随机变量定义

- **定义 (随机变量)** X 是定义在样本空间 Ω 上的实值函数: 对每一个样本点 ω, $X(\omega)$ 是一个实数.
- (更严格的数学定义还要求关于 X 落入区间是事件)。
- 通常将随机变量 $X(\omega)$ 简记为 X.
- 在概率论和数理统计学中, 人们习惯用大写的 X, Y, Z, ξ, η 等表示随机变量. 不够时还可以用 X_1, X_2, $\cdots$ 等表示.

随机变量的事件

- 我们用 $\{X \le x\}$, 或更简单地用 $X \le x$ 表示事件

  $$\{\ \omega \mid X(\omega) \le x\}$$

  .
- 对于实数的集合 A, 我们用 $\{X \in A\}$, 或更简单地用 $X \in A$ 表示事件 $\{\omega \,|\, X(\omega) \in A\}$.
- 于是

  $$\begin{aligned}\{X \in A\} =& \{\omega \mid X(\omega) \in A\},\\ \{a < X \le b\} =& \{\omega \mid a < X(\omega) \le b\}.\end{aligned}$$

- **注:** 这里和以后所述的数集都是高等数学中的实数的 (可测) 集合, 并且对数集 A, 承认 $\{X \in A\}$ 是事件.

例 1.3,1.4: 随机变量的函数

- **例 1.3** 掷一个骰子, 用 X 表示掷出的点数, 则 X, X^2, $X+\sqrt{X}$ 都是样本空间上的函数, 因而都是随机变量.
- **例 1.4** 掷 n 个骰子, 用 Y 表示掷出的点数之和, 则 Y 是随机变量. 对函数 $g(x)$, $X=g(Y)$ 也是随机变量, 因为 $X(\omega)=g(Y(\omega))$ 也是样本空间 Ω 上的函数.

例 1.5: 随机变量与概率

- **例 1.5** 在 52 张扑克中任取 13 张, 求这 13 张牌中恰有 5 张草花的概率.
- **解** 用 X 表示这 13 张牌中草花的张数, 则 $X=5$ 是关心的事件, 容易得到
$$P(X=5)=\frac{C_{13}^5\,C_{39}^8}{C_{52}^{13}}=0.1247.$$
- **注:** 在许多实际问题中, 一个随机变量 X 的含义是十分清楚的, 所以一般不再关心随机变量 X 在样本空间 Ω 上是如何定义的. 可以认为 X 的所有取值就是我们的样本空间。只是在必要的时候才将自变元 ω 写出来.

2.2 离散型随机变量

离散型随机变量

- 有些变量只能取有限个或可列个值，比如，被访问者的性别、年龄、职业，一批产品中次品个数，一个医学试样中白细胞个数，掷两个骰子第一次得到 12 点的时间，等等。
- 另外的变量可以取到区间内任何值，比如温度、气压、长度、时间等测量值。
- **定义 2.1** 如果随机变量 X 只取有限个值 $x_1,x_2,\cdots,x_n$, 或可列个值 $x_1,x_2,\cdots$, 就称 X 是**离散型随机变量**, 简称为**离散随机变量** (discrete random variable).
- 以下就 X 取可列个值的情况加以表述, 对于 X 取有限个值的情况可类似的表述.

分布列

- **定义 2.2** 设 X 是离散随机变量, 称

$$P(X = x_k) = p_k, \quad k \geq 1, \tag{2.1}$$

为 X 的**概率分布** (probability distribution). 称 $\{p_k\}$ 是**概率分布列**, 简称为**分布列**. (distribution sequence).

- 设函数 $f(x)$ 取值于 $\{x_1, x_2, \dots\}$, $f(x_k) = p_k$, 称 $f(x)$ 为随机变量 X 的概率质量函数 (PMF, probability mass function)。

- 当分布列 $\{p_k\}$ 的规律性不够明显时, 也常常用如下的方式表达概率分布,

$$\begin{array}{c|cccc} X & x_1 & x_2 & x_3 & \cdots \\ \hline P & p_1 & p_2 & p_3 & \cdots \end{array} \tag{2.2}$$

- 分布列 $\{p_k\}$ 有如下的性质:
  - (a) $p_k \geq 0$,
  - (b) $\sum_{j=1}^{\infty} p_j = 1$.

- 由于 p_k 是概率, 所以是非负的.

- 下面证明 (b). 对 $k \neq j$, $\{X = x_j\}$ 发生, $\{X = x_k\}$ 就不能发生, 所以 $\{X = x_j\}, j = 1, 2, \cdots$, 互不相容. 利用

$$\Omega = \bigcup_{j=1}^{\infty} \{X = x_j\}.$$

和概率的可列可加性得到

$$1 = P(\Omega) = \sum_{j=1}^{\infty} P(X = x_j) = \sum_{j=1}^{\infty} p_j.$$

两点分布

- **两点分布 (Bernoulli 分布)** $B(1, p)$: 如果 X 只取值 0 或 1, 概率分布是

$$P(X = 1) = p, \quad P(X = 0) = q, \quad p + q = 1, \tag{2.3}$$

就称 X 服从两点分布, 记做 $X \sim \mathrm{B}(1, p)$ 或 $X \sim \mathrm{b}(1, p)$。

- 任何试验, 当只考虑成功与否时, 就可以用两点分布的随机变量描述:

$$X = \begin{cases} 1, & \text{试验成功}, \\ 0, & \text{试验不成功}. \end{cases}$$

二项分布

- **二项分布 (Binomial 分布)**$B(n,p)$**:** 如果随机变量有如下的概率分布:

$$P(X=k) = C_n^k p^k q^{n-k},\ k=0,1,\cdots,n, \quad (2.4)$$
$$(\text{其中 } pq>0,\ p+q=1)$$

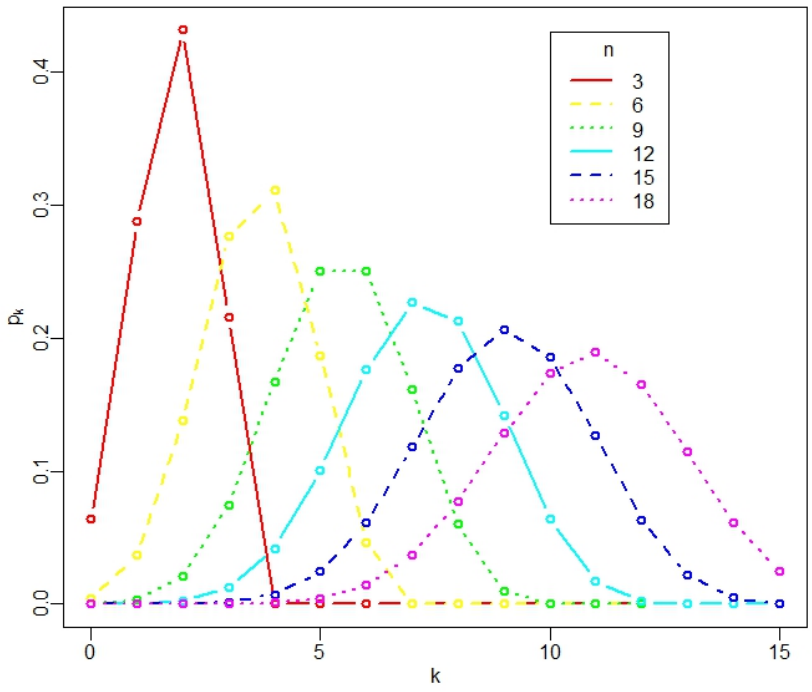
  就称 X 服从二项分布, 记做 $X \sim \mathrm{B}(n,p)$.

- 称为二项分布的原因是 $C_n^k p^k q^{n-k}$ 为二项展开式:

$$(p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k}$$

  的第 $k+1$ 项. B 表示 Binomial.

二项分布的折线图



二项分布的背景

- **二项分布的背景** 设试验 S 成功的概率为 p, 将试验 S 重复 n 次, 用 X 表示成功的次数, 则 $X \sim B(n,p)$.
- 解释: 用 A_j 表示第 j 次试验成功, 则 A_1, A_2, $\cdots$, A_n 相互独立, 且 $P(A_j)=p$.
- 从 n 次试验中选定 k 次试验的方法共有 C_n^k 种. 对第 j 种选法为 $\{j_1,j_2,\cdots,j_k\}$ 成功，其余失败, 用

  $$B_j = A_{j_1}A_{j_2}\cdots A_{j_k}\overline{A}_{j_{k+1}}\overline{A}_{j_{k+2}}\cdots\overline{A}_{j_n}$$

  表示, 则 $\{B_j\}$ 互不相容, 并且

  $$\{X=k\} = \bigcup_{j=1}^{C_n^k} B_j, \quad P(B_j) = p^k q^{n-k}.$$

- 用概率的有限可加性得到

  $$P(X=k) = \sum_{j=1}^{C_n^k} P(B_j) = C_n^k p^k q^{n-k}.$$

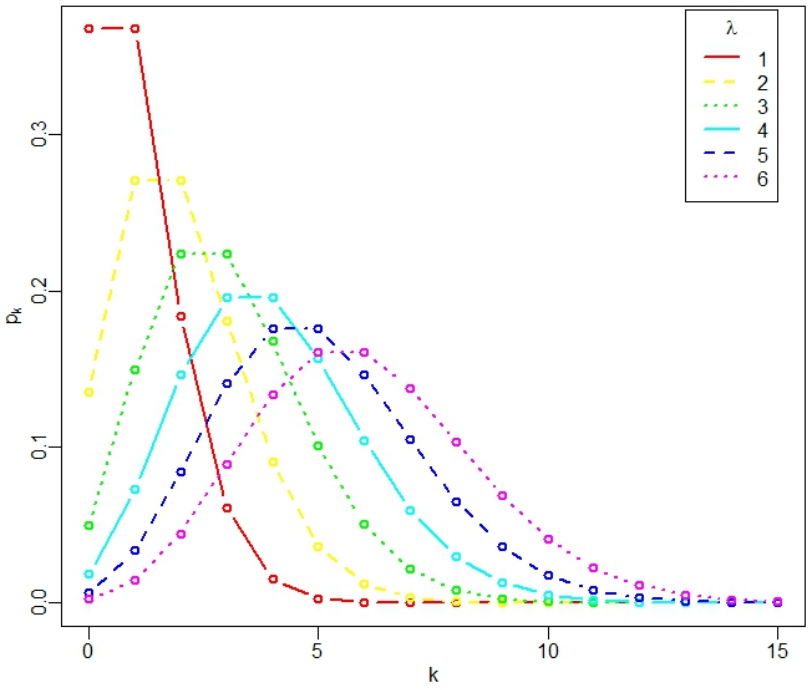
- **泊松分布 (Poisson 分布) Poisson**(λ)**:** 如果随机变量 X 有如下的概率分布:

  $$P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0,1,\cdots, \tag{2.5}$$

  就称 X 服从参数是 λ 的 Poisson 分布, 简记为 $X \sim$ Poisson (λ). 这里 λ 是正常数.

- Poisson 分布的例子：
  - 单位时间放射性粒子个数；
  - 某段高速公路上一年的事故数；
  - 某商场一天中顾客到来个数；
  - 一段时间内接到的电话个数；
  - 等等。

泊松分布的折线图



例 2.1 放射性粒子数

- **例 2.1** 1910 年, 著名科学家 Rutherford(罗瑟福) 和 Geiger(盖克) 观察了放射性物质钋（读 po1）(polonium) 放射 α 粒子的情况.

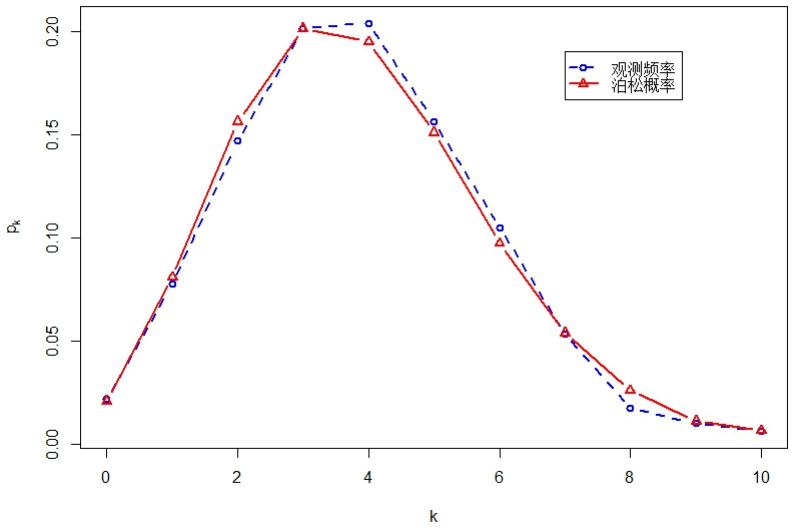
- 他们进行了 $N = 2608$ 次观测, 每次观测 7.5 秒, 一共观测到 10094 个 α 粒子放出, 下面的表 2.2.1 是观测记录. 其中的 Y 是服从 Poisson(3.87) 分布的随机变量, $3.87 = 10094/2608$ 是 7.5 秒中放射出 α 粒子的平均数.

- 用 Y 表示这块放射性钋在 7.5 秒内放射出的 α 粒子数, 表的最后两列表明事件 $\{Y = k\}$ 在 $N = 2608$ 次重复观测中发生的频率和 $P(Y = k)$ 基本相同.

-

观测到的α粒子数k	观测到k个粒子的次数m_k	发生的频率m_k/N	$P(Y=k)$ $Y\sim$ Poisson(3.87)
0	57	0.022	0.021
1	203	0.078	0.081
2	383	0.147	0.156
3	525	0.201	0.201
4	532	0.204	0.195
5	408	0.156	0.151
6	273	0.105	0.097
7	139	0.053	0.054
8	45	0.017	0.026
9	27	0.010	0.011
10+	16	0.006	0.007
总计	2608	0.999	1.00

放射粒子数的观测频率与泊松概率对比图



例 2.2 战争数

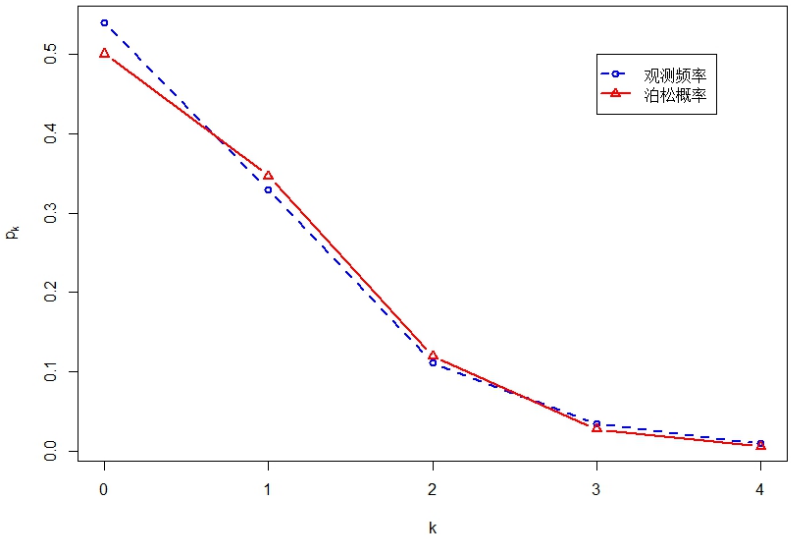
- **例 2.2** 自 1500 至 1931 年的 $N=432$ 年间, 比较重要的战争在全世界共发生了 299 次.

- 以每年为一个时间段的记录见下面。

爆发的战争数k	爆发k次战争的年数m_k	频率m_k/N	$P(Y=k)$
0	223	0.516	0.502
1	142	0.329	0.346
2	48	0.111	0.119
3	15	0.035	0.028
4+	4	0.009	0.005
总计	432	1.000	1.000

- 平均每年发生战争次数 $\lambda = 299/432 \approx 0.69$。
- 上面, $Y \sim \text{Poisson}(0.69)$.

一年中发生战争次数的频率与泊松概率对比图



例 2.3 出租车遇红灯数

- **例 2.3** 设一辆出租车一天内穿过的路口数 Y 服从泊松分布 $\text{Poisson}(\lambda)$, 设各个路口的红绿灯是独立工作的, 在每个路口遇到红灯的概率是 $p(>0)$.
- (1) 已知一辆出租车一天内路过了 k 个路口, 求遇到的红灯数的分布;
- (2) 求一辆出租车一天内遇到的红灯数的分布.
- **解** 设这辆出租车路过的路口数是 Y, 遇到的红灯数是 X. 每到一个路口相当于作一次试验, 遇到红灯是试验成功.

- (1)

$$P(X=m|Y=k)=C_k^m p^m q^{k-m},\quad m=0,1,\ldots,k,\ q=1-p,$$

- (2) $\{Y=j\};\ j=0,1,2,\ldots,$ 构成完备事件组. 利用全概率公式得到

$$\begin{aligned}P(X=m)&=\sum_{k=m}^{\infty}P(Y=k)P(X=m|Y=k)\\&=\sum_{k=m}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}C_k^m p^m q^{k-m}\\&=\sum_{k=m}^{\infty}\frac{(\lambda q)^{k-m}}{m!(k-m)!}e^{-\lambda}(\lambda p)^m\\&=\frac{(\lambda p)^m}{m!}e^{-\lambda}\sum_{j=0}^{\infty}\frac{(\lambda q)^j}{j!}\\&=\frac{(\lambda p)^m}{m!}e^{-\lambda}e^{\lambda q}\\&=\frac{(\lambda p)^m}{m!}e^{-\lambda p},\quad m=0,1,\cdots.\end{aligned}$$

说明 X 服从泊松分布 Poisson$(p\lambda)$.

超几何分布 H(n,M,N)

- 如果 X 的概率分布是

$$P(X=m)=\frac{C_M^m\ C_{N-M}^{n-m}}{C_N^n},\quad m=0,1,\ldots,M, \tag{2.8}$$

就称 X 服从超几何分布, 记作 H(n,M,N).

- 注意对 $k<0$ 或 $k>n$, 约定 $C_n^k=0$. 名称超几何分布来自超几何函数, 类似于二项分布来自二项展开式.

- **例 2.4** N 件产品中恰有 M 件次品, 从中任取 n 件, 用 X 表示这 n 件中的次品数, 则 X 服从超几何分布 (2.8).

- 如果这批产品充分多, 无放回的抽取和有放回的抽取就没有本质的差异:

$$P(X=m)=\frac{C_M^m\ C_{N-M}^{n-m}}{C_N^n}\approx C_n^m p^m(1-p)^{n-m},\quad (p=\frac{M}{N})$$

几何分布

- 如果随机变量 X 有如下的分布

$$\begin{aligned} &P(X=k)=q^{k-1}p, \quad k=1,2,\cdots, \\ &pq>0,\ p+q=1, \end{aligned} \tag{2.10}$$

  就称 X 服从参数是 p 的几何分布.

- 设某试验成功概率为 p，独立地重复此试验直到第一次成功，则第一次成功需要的试验次数分布为参数 p 的几何分布。

- **例 2.5** 甲向一个目标射击, 直到击中为止. 用 X 表示首次击中目标时的射击次数. 如果甲每次击中目标的概率是 p, 则 X 服从几何分布 (2.10).

- **解** 用 A_j 表示甲第 j 次没击中目标, 由 $\{A_j\}$ 的独立性得到

$$\begin{aligned} P(X=k)=&P(A_1A_2\cdots A_{k-1}\overline{A}_k) \\ =&q^{k-1}p, \quad k=1,2,\cdots. \end{aligned} \tag{2.11}$$

2.3 连续型随机变量

连续型随机变量定义

- 在线段上随机投点的位置，温度、气压、电压、电流等物理量等等，理论上可以在取到某个区间任何实数值。这样取值的随机变量称为连续型随机变量。

- **定义 3.1** 设 X 是随机变量, 如果存在非负函数 $f(x)$ 使得对任何满足 $-\infty \le a < b \le \infty$ 的 a, b, 有

$$P(a < X \le b) = \int_a^b f(x)\,dx, \tag{3.1}$$

  就称 X 是连续型随机变量, 称 $f(x)$ 是 X 的**概率密度函数**, 简称为**概率密度** (probability density) 或**密度**.

分布密度性质

- 设 $f(x)$ 是 X 的概率密度, 则 $f(x)$ 有如下的基本性质.

  (a) $\int_{-\infty}^{\infty} f(x)\,dx = 1$,

(b) $P(X=a)=0$. 于是 $P(a<X\le b)=P(a\le X\le b)$,

(c) 对数集 A(严格意义下要求可测性),

$$P(X\in A)=\int_A f(x)\,dx. \tag{3.2}$$

- **证明:** (a) 由

$$\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X \le \infty) = 1,$$

可得。

- (b)

$$\Pr(X=a)\le \Pr(X\in(a-\varepsilon,a])=\int_{a-\varepsilon}^{a} f(x)\,dx\to 0,\quad \varepsilon\to 0.$$

- (c) 不证。

概率密度的意义

- 概率密度与离散型随机变量的分布列有很大差别，分布列 $p_k=\Pr(X=x_k)$ 本身就是 X 取 x_k 的概率;
- 连续型随机变量取任何一个特定值的概率都等于零; $f(x)$ 是一个相对的概念，如果 $f(x_2)=2f(x_1)$，可以认为 X 在 x_2 “附近”取值的概率比 X 在 x_1 附近取值的概率大一倍, 严格讲，假设 $f(x)$ 在 x_1 和 x_2 处连续，

$$\frac{\Pr(x_2-\varepsilon<X\le x_2+\varepsilon)}{\Pr(x_1-\varepsilon<X\le x_1+\varepsilon)}=\frac{\int_{x_2-\varepsilon}^{x_2+\varepsilon} f(x)\,dx}{\int_{x_1-\varepsilon}^{x_1+\varepsilon} f(x)\,dx}\to\frac{f(x_2)}{f(x_1)}=2$$

均匀分布

- **均匀分布 (Uniform 分布)** $\mathrm{U}(a,b)$: 对 $a<b$, 如果 X 的密度是

$$f(x)=\begin{cases}\frac{1}{b-a}, & x\in(a,b),\\ 0, & x\notin(a,b).\end{cases} \tag{3.3}$$

就称 X 服从区间 (a,b) 上的均匀分布, 记做 $X\sim\mathrm{U}(a,b)$. 这里 U 是 Uniform 的缩写.

- 明显, 表达式 (3.3) 中的区间 (a,b) 也可以写成 $(a,b]$, $[a,b)$ 或 $[a,b]$.

- 采用 (a,b) 的**示性函数** (indicator function)

$$\mathrm{I}_{(a,b)} = \begin{cases} 1, & x \in (a,b), \\ 0, & x \notin (a,b), \end{cases}$$

还可以将 (3.3) 中的密度 $f(x)$ 写成 $f(x) = \frac{1}{b-a}\mathrm{I}_{(a,b)}$.

例 3.1 等车

- **例 3.1** 每天的整点 (如 9 点, 10 点, 11 点等) 甲站都有列车发往乙站. 一位要去乙站的乘客在 9 点至 10 点之间随机到达甲站. 用 Y 表示他的等车时间, 计算他候车时间小于 30 分钟的概率.

- **解** 题目中随机到达的含义指在等长的时间段中到达的可能性相同.

- 用 X 表示他的到达时刻, X 在 0 至 60 分钟内均匀分布, 有密度函数

$$f(x) = \frac{1}{60}\mathrm{I}_{(0,60)}.$$

- $\{Y < 30\text{分钟}\}$ 表示该乘客在 9 : 30 至 10 : 00 之间到达, 这是和 $\{30 < X \le 60\}$ 等价的, 于是

$$P(Y < 30\text{分钟}) = P(30 < X \le 60) = \int_{30}^{60} f(x)dx = \frac{1}{2}.$$

- **指数分布 (Exponential 分布)** $\mathrm{E}(\lambda)$: 对正常数 λ, 如果 X 的密度是

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0, \end{cases} \tag{3.4}$$

就称 X 服从参数 λ 的指数分布, 记做 $X \sim \mathrm{E}(\lambda)$. 这里 E 是 Exponential 的缩写.

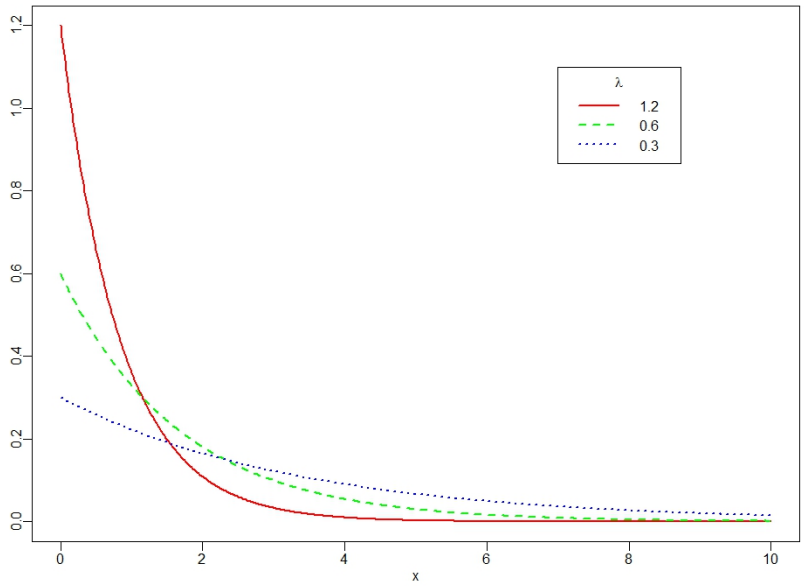
- 通常还把 (3.4) 简记为

$$f(x) = \lambda e^{-\lambda x}, \quad x \ge 0,$$

或

$$f(x) = \lambda e^{-\lambda x}\mathrm{I}_{[0,\infty)}.$$

指数分布的密度图示例



指数分布的无后效性

- 指数分布经常用来表示电子元件寿命、事件到来间隔时间等。这样的量经常具有“无后效性”，即已经存活 (等待) 了多长时间对还会再存活 (等待) 多长时间没有影响。
- **非负随机变量**: 若随机变量 X 满足 $\Pr(X<0)=0$ 则称 X 为非负随机变量。
- **定理 3.1** 设 X 是连续型非负随机变量, 则 X 服从指数分布的充分必要条件是对任何 $s,t\geq 0$, 有

$$P(X>s+t|X>s)=P(X>t). \tag{3.5}$$

- 性质 (3.5) 称为**无后效性**. 无后效性是指数分布的特征.
- 如果 X 表示某仪器的工作寿命, 无后效性 (3.5) 的解释是: 当仪器工作了 s 小时后再能继续工作 t 小时的概率等于该仪器刚开始就能工作 t 小时的概率. 说明该仪器的使用寿命不随使用时间的增加发生变化, 或说仪器是“永葆青春”的.
- 一般来说, 电子元件和计算机软件等具备这种性质, 它们本身的老化是可以忽略不计的, 造成损坏的原因是意外的高电压, 计算机病毒等等. 青花盘的使用寿命也可被认为有无后效性.

例 3.2 粒子到来间隔时间

- 设时间 $(0,t]$ 内有 $N(t)$ 个粒子放射出来。$N(t) \sim \text{Poisson}(\mu t)$。
- 设 X 为第一个粒子发射出来的时刻，则

$$\{X > t\} = \{N(t) = 0\}$$

- $$\Pr(X > t) = \Pr(N(t) = 0) = e^{-\mu t}\frac{(\mu t)^0}{0!} = e^{-\mu t}$$
- 对任何 $0 \le a < b$ 有

$$\begin{aligned} &P(a < X_1 \le b) = P(X_1 > a) - P(X_1 > b) \\ =&e^{-\mu a} - e^{-\mu b} = \int_a^b \mu e^{-\mu x}\, dx. \end{aligned}$$

- 即 X 的概率密度为 $f(x) = \mu e^{-\mu x} I_{[0,\infty)}(x)$.

正态分布

- **正态分布 (Normal 分布)** $\text{N}(\mu, \sigma^2)$: 设 μ 是常数, σ 是正常数. 如果 X 的密度是

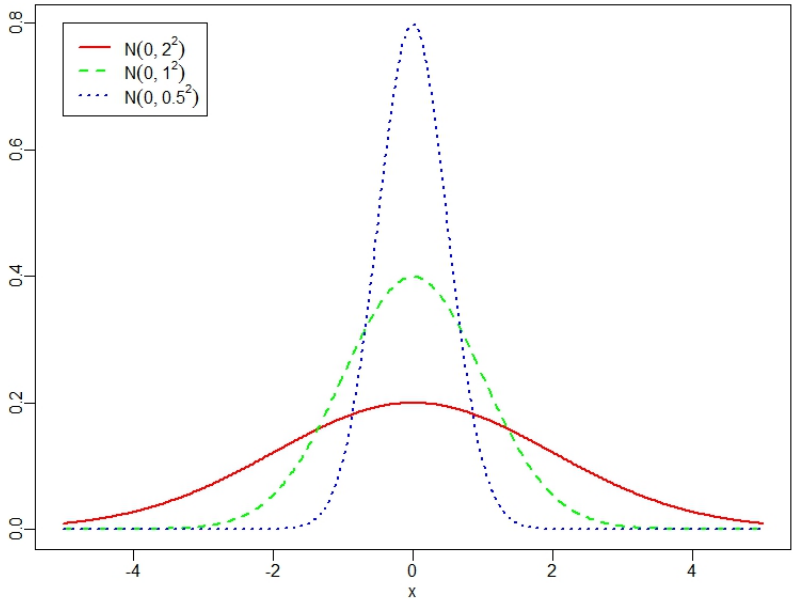
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in R, \tag{3.8}$$

  就称 X 服从参数为 (μ, σ^2) 的正态分布, 记做 $X \sim \text{N}(\mu, \sigma^2)$. 这里 N 是 Normal 的缩写.

- 特别, 当 $X \sim \text{N}(0,1)$ 时, 称 X 服从**标准正态分布** (standard normal distribution). 标准正态分布的密度函数有特殊的地位, 所以用一个特定的符号 φ 表示:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}. \tag{3.9}$$

正态分布的密度图示例



正态分布密度特点

- 参数 μ 是密度的中心和最大值点，密度在 μ 两侧对称;
- 参数 σ 代表了密度的宽度，σ 越大密度越宽。(见演示)
- 正态分布的随机变量 X 具有大部分值靠近 μ 的特点 (经验规则):

$$\begin{aligned}\Pr(|X-\mu|\le\sigma) &=68.27\%\\ \Pr(|X-\mu|\le 2\sigma) &=95.45\%\\ \Pr(|X-\mu|\le 3\sigma) &=99.73\%\\ \Pr(|X-\mu|>6\sigma) &=1.96\times 10^{-9}\end{aligned}$$

  出现在 $\mu\pm k\sigma$(k=2,3,6 等) 外的点认为是比较值得注意的点。

- 记 $\Phi(x)=\int_{-\infty}^{x}\varphi(t)\,dt$, $\Phi(x)$ 有表格。另外 $\Phi(-x)=1-\Phi(x)$。
- 对 $X\sim \mathrm{N}(\mu,\sigma^2)$,

$$\Pr(X\in(a,b])=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$$

正态分布的历史

- 正态分布最早由 Gauss 在研究测量误差时得到, 所以正态分布又被称为 Gauss 分布.

- 在布朗运动的研究中, 人们也得到了正态分布.
- 正态分布在概率论和数理统计中有特殊的重要地位. 事实表明, 产品的许多质量指标, 生物和动物的许多生理指标等都服从或近似服从正态分布.
- 大量相互独立且有相同分布的随机变量的累积也近似服从正态分布 (参考二项分布当 n 较大时的概率分布图形).

例 3.3 零件长度

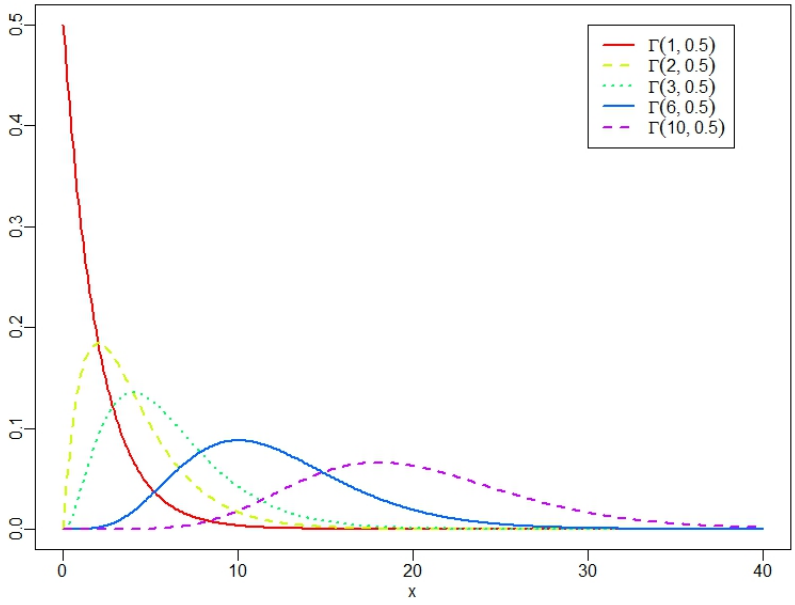
- **例 3.3** 一台机床加工的部件长度服从正态分布 $N(10, 36\times10^{-6})$. 当部件的长度在 10 ± 0.01 内为合格品, 求一部件是合格品的概率.
- **解** 用 X 表示生产的一个部件的长度, 则 $X\sim N(10, 36\times10^{-6})$. 事件
$$\{-0.01\le X-10\le 0.01\}$$
表示这个部件是合格品.
- $$\begin{aligned}&P(-0.01\le X-10\le 0.01)\\ =&P(\frac{-0.01}{6\times10^{-3}}\le\frac{X-10}{6\times10^{-3}}\le\frac{0.01}{6\times10^{-3}})\\ =&\Phi(1.67)-\Phi(-1.67)=2\Phi(1.67)-1\\ =&2\times0.9525-1=0.905.\end{aligned}$$
- 这个概率令人太不满意了, 说明这台机床的质量有问题. 以后会知道质量问题可以由参数 σ^2 体现出来 (参考习题 2.17).

Gamma 分布

- **Gamma 分布** $\Gamma(\alpha,\lambda)$: 设 α,λ 是正常数, $\Gamma(\alpha)$ 由积分
$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,dx \tag{3.14}$$
定义. 如果 X 的密度是
$$f(x)=\begin{cases}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x\ge0,\\ 0, & x<0.\end{cases} \tag{3.15}$$
就称 X 服从参数 (α,λ) 的 Gamma 分布, 记做 $X\sim\Gamma(\alpha,\lambda)$. 这里 Γ 是 Gamma 的简写.

- 注意 $\alpha = 1$ 时 $\Gamma(1, \lambda)$ 即 $\mathrm{E}(\lambda)$。

伽马分布的密度图示例



Gamma 分布的历史

- 英国著名统计学家 Pearson 在研究物理, 生物及经济中的随机变量时, 发现很多连续型随机变量的分布都不是正态分布.
- 这些随机变量的特点是只取非负值, 于是他致力于这类随机变量的研究.
- 从 1895 年至 1916 年间, Pearson 连续发表了一系列的连续分布密度曲线, 认为这些曲线可以包括常见的单峰分布, 其中就有 Gamma 分布.
- 在气象学中, 干旱地区的年、季或月降水量被认为服从 Γ 分布, 指定时间段内的最大风速等也被认为服从 Γ 分布.

例 3.4: 指数分布随机变量的和

- **例 3.4** 在 Poisson 分布的例子中, 用 S_k 表示从开始到观测到第 k 个 α 粒子的时间, 可以证明 S_k 服从 $\Gamma(k, \lambda)$ 分布.

2.4 概率分布函数

2.4.1 概率分布函数

概率分布函数

- 为了计算事件 $\{X \in (a,b]\}$ 的概率，如果 X 是离散型随机变量，则

$$\Pr(X \in (a,b]) = \sum_{x_k \in (a,b]} p_k$$

- 如果 X 是连续型随机变量，则

$$\Pr(X \in (a,b]) = \int_a^b f(x)\,dx$$

- 事实上，如果我们定义 $F(x) = \Pr(X \le x)$，则 $\Pr(X \in (a,b]) = F(b) - F(a)$。这样的 $F(x)$ 可以帮助我们计算 $\{X \in (a,b]\}$ 的概率。

概率分布函数定义

- **定义 4.1** 对随机变量 X, 称 x 的函数

$$F(x) = P(X \le x), \quad -\infty \le x \le \infty, \tag{4.1}$$

为 X 的**概率分布函数**, 简称为**分布函数** (distribution function), 也称为累积 (cumulative) 分布函数。

- **例 4.1** $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt$ 是标准正态分布的分布函数.

离散型随机变量的分布函数

- 从定义看出, 如果 X 是离散型随机变量, 有概率分布

$$p_k = P(X = x_k),\ k = 1, 2, \cdots, \tag{4.2}$$

则 X 的分布函数

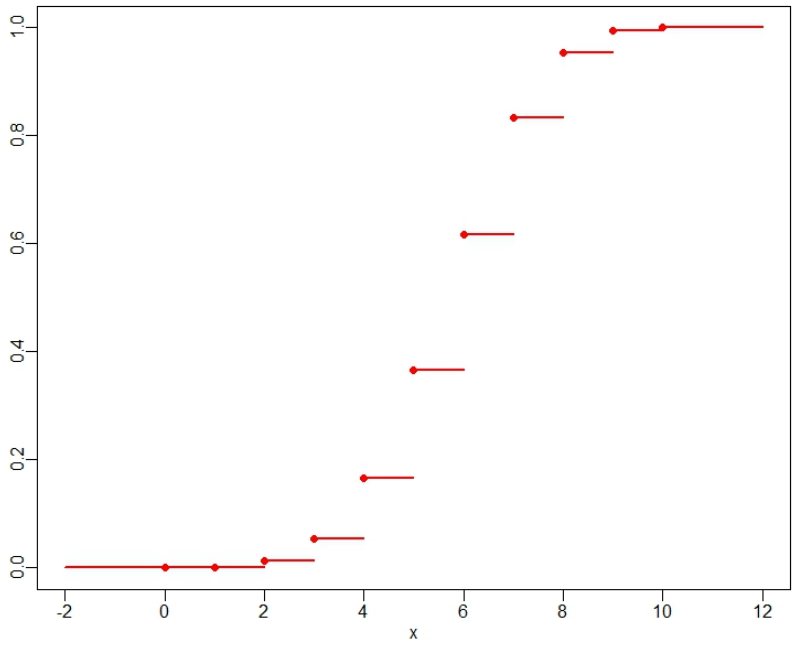
-

$$F(x) = P(X \le x) = P(\bigcup_{j:\ x_j \le x} \{X = x_j\}) = \sum_{j:\ x_j \le x} p_j \tag{4.3}$$

- 是单调不减的阶梯函数.

- 它在每个 x_j 有跳跃 p_j.
- 这时, 我们也称 $F(x)$ 是分布列 $\{p_j\}$ 的分布函数.
- 见二项分布 $B(10, 0.6)$ 的分布函数图, 横坐标是 x, 纵坐标是 $F(x)$. 从图形可以看出, $F(x)$ 是单调不减右连续函数.

B(10,0.6) 的分布函数图



连续型随机变量的分布函数

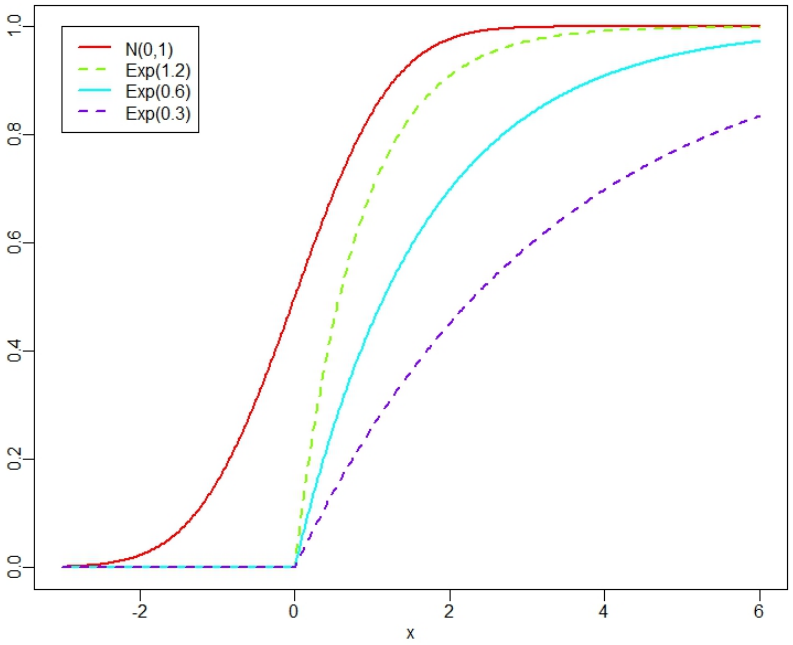
- 如果 X 是连续型随机变量, 有概率密度 $f(x)$, 则

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \tag{4.4}$$

是连续函数, 并且在 $f(x)$ 的连续点 x 有 $f(x) = F'(x)$. 我们也称 $F(x)$ 是 $f(x)$ 分布函数.

- 见图形。

N(0,1), $\mathcal{E}$(1.2), $\mathcal{E}$(0.6), $\mathcal{E}$(0.3) 分布函数图



分布函数性质

- 分布函数 $F(x)$ 的常用性质:

  (1) F 单调不减右连续,

  (2) $F(\infty)=1$, $F(-\infty)=0$.

- **证明** (1) 对 $x<y$, 单调不减性由

$$\{x<X\le y\}=\{X\le y\}-\{X\le x\}$$

  和 $P(x<X\le y)=P(X\le y)-P(X\le x)=F(y)-F(x)\ge 0$ 得到.

- 由于 n 越大, 集合 $\{X\le x+1/n\}$ 越小, 所以用 F 的单调性和概率 P 的连续性得到

$$\begin{aligned}\lim_{\delta\downarrow 0}F(x+\delta)&=\lim_{n\to\infty}F(x+1/n)\\&=\lim_{n\to\infty}P(X\le x+1/n)\\&=P(\cap_{n=1}^{\infty}\{X\le x+1/n\})\\&=P(X\le x)=F(x).\end{aligned}$$

- (2) 由 $F(\infty)=P(X\le\infty)=P(\Omega)=1$ 和 $F(-\infty)=P(X\le-\infty)=P(\emptyset)=0$ 得到 (2).

密度与分布函数

- 对于连续型的随机变量, 密度函数通过 (4.4) 唯一决定分布函数.
- 反过来，如果 X 的分布函数 $F(x)$ 在 $(-\infty,\infty)$ 有连续的导函数，则 $F'(x)$ 是 X 的密度函数。
- 对于区间 (a,b)，如果 $F(x)$ 在 (a,b) 内有连续的导函数，且 $F(a)=0$, $\lim_{x\to b+0}F(x)=F(b)=1$ (这两个条件相当于 $P(X<=a)=0$, $P(X<b)=1$，即 $P(a<X<b)=1$)，则 $F'(x), x\in(a,b)$ 是 X 的密度函数。
- 更一般地，如果 $F(x)$ 连续且除去有限个点之外都连续可导，则 $F'(x)$ 是 X 的密度函数。

密度与分布函数 (续)

- **定理 4.1** 设 X 的分布函数 F 连续, 数集 A 中任何两点之间的距离大于正数 δ. 如果在 A 外导数 $F'(x)$ 存在且连续, 则

$$f(x)=\begin{cases}F'(x), & \text{当 } x\notin A,\\ 0, & \text{当 } x\in A\end{cases} \tag{4.5}$$

  是 X 的密度函数.
- 定理条件中例外集 A 一定是有限或可列的。
- 连续型分布的分布函数一定是连续的，分布函数如果不连续就不是连续型分布。
- 除了连续型分布和离散型分布以外还存在其它类型的分布。如：零过多数据的分布。

2.4.2 常见分布的分布函数

均匀分布的分布函数

- 若 $X\sim \mathrm{U}(0,1)$，则其分布密度为

$$f(x)=1,\quad x\in(0,1)$$

- 其分布函数为

$$F(x)=\begin{cases}0 & x\le 0\\ \int_0^x 1\cdot dt=x, & x\in(0,1)\\ 1 & x\ge 1\end{cases}$$

- 若 $X\sim \mathrm{U}(a,b)$，则其分布密度为

$$f(x)=\frac{1}{b-a},\quad x\in(a,b)$$

- 其分布函数为

$$F(x)=\begin{cases}0 & x\le a\\ \int_a^x \frac{1}{b-a}\cdot dt=\frac{x-a}{b-a}, & x\in(a,b)\\ 1 & x\ge b\end{cases}$$

正态分布的分布函数

- 设 $X\sim \mathrm{N}(0,1)$，则 X 的分布函数为

$$\Phi(x)=\int_{-\infty}^{x}\varphi(t)\,dt$$

其中

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

- 设 $X\sim \mathrm{N}(\mu,\sigma^2)$，下一节将证明其分布函数为

$$F(x)=\Phi(\frac{x-\mu}{\sigma})$$

指数分布的分布函数

- 若 $X\sim \mathrm{E}(\lambda)$, 则其分布密度为

$$f(x)=\lambda e^{-\lambda x},\quad x\ge 0$$

- 其分布函数为

$$F(x)=\int_0^x \lambda e^{-\lambda t}\,dt=1-e^{-\lambda x},\quad x\ge 0$$

Gamma 分布的分布函数

- 若 $X \sim \Gamma(\alpha, \lambda)$，则其分布密度为

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0$$

- 当 $\lambda = 1$ 时其分布函数为

$$I_\alpha(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t}\, dt, t \geq 0$$

  称为不完全 Gamma 函数，是没有解析表达式的。

- 对一般 $\Gamma(\alpha, \lambda)$ 其分布函数为

$$F(x) = I_\alpha(\lambda x), \quad x \geq 0$$

2.5 随机变量函数的分布

随机变量函数分布—例 5.1

- **例 5.1** 设 X 有如下的概率分布

X	-2	-1	0	1	3
P	0.1	0.2	0.3	0.2	0.2

  求 $Y = X^2$ 的分布.

- **解** Y 的取值是 $0, 1, 4, 9$, 而且

$$\begin{aligned}
&P(Y = 0) = P(X = 0) = 0.3;\\
&P(Y = 1) = P(|X| = 1) = 0.2 + 0.2 = 0.4;\\
&P(Y = 4) = P(|X| = 2) = P(X = -2) = 0.1;\\
&P(Y = 9) = P(X = 3) = 0.2.
\end{aligned}$$

- 于是 Y 有分布

Y	0	1	4	9
P	0.3	0.4	0.1	0.2

例 5.2 均匀分布的反函数

- **例 5.2** 设 $X \sim \mathrm{U}(0,1)$, $\Phi^{-1}(p)(p \in (0,1))$ 是 $\Phi(x)$ 的反函数, 求 $Y = \Phi^{-1}(X)$ 的分布.

- **解**

$$F_Y(y) = P(\Phi^{-1}(X) \le y) = P(X \le \Phi(y)) = \Phi(y),\ \forall x \in \mathbb{R},$$

所以 $Y \sim N(0,1)$.

例 5.3: 正态分布

- **例 5.3** 设 $X \sim N(\mu, \sigma^2)$, 则 $Y = (X-\mu)/\sigma$ 服从标准正态分布 $N(0,1)$, 且 X 的分布函数为 $\Phi(\frac{x-\mu}{\sigma})$.

- **解** 先求 Y 的分布函数 $F_Y(y)$. 设 $F_X(x)$ 是 X 的分布函数, 则 F_X 连续可导, 并且有

$$F'_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big).$$

- 于是,

$$\begin{aligned} F_Y(y) =& P(Y \le y) = P((X-\mu)/\sigma \le y) \\ =& P(X \le y\sigma + \mu) = F_X(y\sigma+\mu) \end{aligned}$$

关于 y 连续可导,

- 对 y 求导数得到概率密度

$$\begin{aligned} f_Y(y) =& F'_Y(y) = F'_X(y\sigma+\mu)\sigma \\ =& \frac{\sigma}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y\sigma+\mu-\mu)^2}{2\sigma^2}\right) \\ =& \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \varphi(y). \end{aligned}$$

说明 $Y \sim N(0,1)$.

- 因为 $F_Y(y) = F_X(\mu+\sigma y)$ 所以 $F_X(x) = F_Y(\frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma})$.

变换

- 设 X 取值于区间 $(a,b)(-\infty \le a < b \le \infty)$, 密度为 $f(x)$。
- 设 $g(x)$ 是 (a,b) 上可导函数且 $g'(x) > 0$，$\forall x \in (a,b)$，则 $g(x)$ 是严格单调增的一一变换，值域也是区间，记值域为 D，$g(x)$ 有逆变换 $h(y)$，$y \in D$，且 $h'(y) = 1/g'(h(y))$。
- 于是

$$\begin{aligned} F_Y(y) =& P(g(X) \le y) = P(X \le h(y)) = F_X(h(y)), \\ F'_Y(y) =& f(h(y))\, h'(y), \end{aligned}$$

- 若 $F'_Y(y)$ 除去有限个点之外连续可导，则 $f_Y(y) = f(h(y))\, h'(y)$。
- 如果 $g(x)$ 是严格单调减的，则

$$\begin{aligned} F_Y(y) =& P(g(X) \le y) = P(X \ge h(y)) \\ =& 1 - P(X < h(y)) = 1 - F_X(h(y)) \\ F'_Y(y) =& -f(h(y))h'(y) = f(h(y))|h'(y)|. \end{aligned}$$

- 下面的定理允许变换不是严格单调的一一变换。

定理

- **定理 5.1** 设 X 有密度函数 $f(x)$, $D \subset \mathbb{R}$, $Y = g(X)$, $P(Y \in D) = 1$. 如果存在函数 $h_i(y)$ 使得

  (1) 对 $y \in D$, $\{Y = y\} = \bigcup_{i=1}^{n} \{X = h_i(y)\}$,

  (2) 每个 $h_i(y)$ 是 D 到其值域 D_i 的可逆映射, 有连续的导数,

  (3) 值域 $D_1, D_2, \cdots, D_n$ 互不相交,

  则 Y 有密度函数

$$f_Y(y) = \begin{cases} \sum_{i=1}^{n} f(h_i(y))|h'_i(y)|, & y \in D, \\ 0, & y \in \overline{D}. \end{cases} \tag{5.1}$$

  证明见附录 A3.

注

- 定理中 D 是 $Y=g(X)$ 的值域，$\cup_{i=1}^{n}D_i$ 是 X 的值域，D_i 是 $h_i(y)$ 的值域，$h_i(y)$ 是变换 $g(x)$ 的多个逆变换中的一个。
- 定理中的 n 也可以是 ∞. 计算随机变量的函数的概率密度的更直接方法请参考附录 D.
- 为了方便记忆, 可以把 (5.1) 写成

$$f_Y(y)=\sum_{i=1}^{n}\Big|\frac{dF_X(h_i(y))}{dy}\Big|,\ y\in D. \tag{5.2}$$

推论

- **推论**: 设随机变量 X 取值于 $C\subset\mathbb{R}$, $Y=g(X)$, $g(x)$ 是 C 到 $D\subset\mathbb{R}$ 的一一变换，$x=h(y)=g^{-1}(y)$ 是 $g(x)$ 的反函数，设 $h(y)$ 有连续的导数。则

$$f_Y(y)=f(h(y))|h'(y)|,\ y\in D.$$

例 5.4: 正态分布的线性变换

- **例 5.4** 设常数 $a\neq 0$, $X\sim N(\mu,\sigma^2)$, 则 $Y=aX+b$ 服从正态分布 $N(a\mu+b,a^2\sigma^2)$. 特别地, $Y=\frac{X-\mu}{\sigma}\sim \mathrm{N}(0,1)$。
- **解** X 有密度函数

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big).$$

- 设 $D=(-\infty,\infty)$, 则 $P(Y\in D)=1$,

$$\{Y=y\}=\{aX+b=y\}=\{X=(y-b)/a\}.$$

  $h(y)=(y-b)/a$ 满足定理 5.1(或推论) 的条件, 有导数 $h'(y)=1/a$.
- 利用定理 5.1 得到 Y 的概率密度

$$\begin{aligned}f_Y(y)=&f_X(h(y))|h'(y)|\\=&\frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{[(y-b)/a-\mu]^2}{2\sigma^2}\Big)\frac{1}{|a|}\\=&\frac{1}{\sqrt{2\pi}\sigma|a|}\exp\Big(-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\Big)\end{aligned}$$

- 再由正态分布的定义知道 $Y\sim N(a\mu+b,a^2\sigma^2)$.

例 5.5: 正态分布的平方

- 设 $X \sim N(0,1)$, 求 $Y = X^2$ 的分布.
- **解** 设 $D = (0, \infty)$, 则 $P(Y \in D) = P(|X| > 0) = 1$.

$$\{Y = y\} = \{X = \sqrt{y}\} + \{X = -\sqrt{y}\},\ y \in D.$$

- $h_1(y) = \sqrt{y}$, $h_2(y) = -\sqrt{y}$ 满足定理 5.1 的条件.
- 由定理 5.1 得到 Y 的密度

$$\begin{aligned} f_Y(y) =& f_X(h_1(y))|h_1'(y)| + f_X(h_2(y))|h_2'(y)| \\ =& \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}h_1^2(y)\Big)\frac{1}{2\sqrt{y}} \\ &+ \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}h_2^2(y)\Big)\frac{1}{2\sqrt{y}} \\ =& \frac{1}{\sqrt{2\pi y}} e^{-y/2},\ \ y \in (0, \infty). \end{aligned}$$

例 5.6

- 设 r 是正数, $X \sim \mathrm{U}(0, 2\pi)$, 求 $Y = r\cos X$ 的密度.
- **解** 设 $D = (-r, r)$, 则 $P(Y \in D) = 1$.
- 对于 $y \in D$, 有

$$\begin{aligned} &\{Y = y\} = \{\cos X = y/r\} \\ =&\{\cos X = y/r, X \in (0,\pi)\} \bigcup \{\cos X = y/r, X \in (\pi, 2\pi)\} \\ =&\{\cos X = y/r, X \in (0,\pi)\} \bigcup \{\cos(2\pi - X) = y/r, X \in (\pi, 2\pi)\} \\ =&\{X = \arccos(y/r)\} \bigcup \{X = 2\pi - \arccos(y/r)\}, \end{aligned}$$

- $h_1(y) = \arccos(y/r)$, $h_2(y) = 2\pi - \arccos(y/r)$ 满足定理 5.1 的条件, 在 D 中有导数

$$h_1'(y) = \frac{1}{\sqrt{r^2 - y^2}},\ h_2'(y) = -\frac{1}{\sqrt{r^2 - y^2}}.$$

- 因为 X 有分布函数

$$F_X(x) = \frac{x}{2\pi},\ x \in (0, 2\pi),$$

- 于是 Y 有密度函数

$$\begin{aligned} f_Y(y) =& f_X(h_1(y))|h_1'(y)| + f_X(h_2(y))|h_2'(y)| \\ =& \frac{1}{\pi\sqrt{r^2-y^2}}, \quad y \in (-r, r). \end{aligned}$$

第三章　随机向量及其分布

3.1　随机向量及其联合分布

随机向量

- 如果 X, Y 都是随机变量, 就称 (X, Y) 是二维随机向量, 简称为**随机向量** (random vector).
- 对于随机向量 (X, Y), 我们称

$$F(x, y) = P(X \leq x, Y \leq y) \tag{1.1}$$

  为 (X, Y) 的**联合概率分布函数**, 简称为**联合分布** (joint distribution).

- 容易证明, 联合分布函数 $F(x, y)$ 是 x 的单调不减函数, 也是 y 的单调不减函数.
- 对随机事件 $A, B, A_1, A_2, \cdots, A_n$, 以后用 $\{A, B\}$ 表示 AB, 用 $\{A_1, A_2, \cdots, A_n\}$ 表示 $\bigcap\limits_{j=1}^{n} A_j$.

边缘分布

- 设 $F(x, y)$ 是 (X, Y) 的联合分布, 则 X,Y 分别有概率分布

$$\begin{aligned} F_X(x) =& P(X \leq x, Y \leq \infty) = F(x, \infty), \\ F_Y(y) =& P(X \leq \infty, Y \leq y) = F(\infty, y). \end{aligned}$$

- 我们称 X 的分布函数 $F_X(x)$, Y 的分布函数 $F_Y(x)$ 为 (X, Y) 的 **边缘分布函数** (marginal distribution function).

独立随机变量

- **定义 1.1** 称随机变量 X, Y 独立, 如果对任何实数 x, y, 事件 $\{X \le x\}$ 和 $\{Y \le y\}$ 独立.
- 按照定义 1.1, X, Y 独立的充分必要条件是对任何 x, y,

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

或等价地有

$$F(x, y) = F_X(x)F_Y(y).$$

- **定义 1.2** 设 $X_1, X_2, \cdots$ 是随机变量,

(1) 如果对任何实数 $x_1, x_2, \cdots, x_n$,

$$\begin{aligned} &P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n) \\ =&P(X_1 \le x_1)P(X_2 \le x_2)\cdots P(X_n \le x_n), \end{aligned}$$

就称随机变量 $X_1, X_2, \cdots, X_n$ 相互独立.

(2) 如果对任何 n, $X_1, X_2, \cdots, X_n$ 相互独立, 就称随机变量的序列 $\{X_j\} = \{X_j : j = 1, 2, \cdots\}$ 相互独立. 这时称 $\{X_j\}$ 是**独立序列** (independent sequence).

独立的性质

- 当 $X_1, X_2, \cdots, X_n$ 是来自相互独立进行的随机试验的随机变量时, 它们相互独立.
- 常数与任何随机变量独立.
- **定理 1.1** 设 $X_1, X_2, \cdots, X_n$ 相互独立, 则有如下的结果.

(1) 对于数集 A_1, A_2, $\cdots$, A_n, 事件

$$\{X_1 \in A_1\},\ \{X_2 \in A_2\},\ \cdots\ \{X_n \in A_n\}$$

相互独立,

(2) 对于一元函数 $g_1(x), g_2(x), \cdots, g_n(x)$, 随机变量 $Y_1 = g_1(X_1), Y_2 = g_2(X_2), \cdots, Y_n = g_n(X_n)$ 相互独立,

(3) 对于 k 元函数 $\varphi(x_1, x_2, \cdots, x_k)$, 随机变量 $\varphi(X_1, X_2, \cdots, X_k)$, X_{k+1}, $X_{k+2}, \cdots$, X_n 相互独立.

- 略去定理证明.

n 元随机向量

- 如果 $X_1, X_2, \ldots, X_n$ 都是随机变量, 就称 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是 n 维随机向量, 也简称为随机向量.
- **定义 1.3** 设 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是随机向量, 称 $\mathbb{R}^n$ 上的 n 元函数

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n) \tag{1.2}$$

为 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 的**联合分布函数**, 简称为**联合分布**.

- 设随机向量 $(X_1, X_2, \ldots, X_n)$ 有联合分布 $F(x_1, x_2, \ldots, x_n)$, X_i 有分布函数 $F_i(x_i)$, 根据随机变量独立性的定义知道, $X_1, X_2, \ldots, X_n$ 相互独立的充分必要条件是对任何 $(x_1, x_2, \ldots, x_n)$, 有

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1)F_2(x_2)\ldots F_n(x_n). \tag{1.3}$$

3.2 离散型随机向量及其分布

二元离散型随机向量

- 如果 X, Y 都是离散型随机变量, 就称 (X, Y) 是离散型随机向量. 设离散型随机向量 (X, Y) 有概率分布

$$p_{i,j} = P(X = x_i, Y = y_j), \quad i, j \ge 1 \tag{2.1}$$

则 X 和 Y 分别有概率分布

$$\begin{aligned} p_i \equiv & P(X = x_i) = \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = \sum_{j=1}^{\infty} p_{i,j}, \ i \ge 1, \\ q_j \equiv & P(Y = y_j) = \sum_{i=1}^{\infty} P(X = x_i, Y = y_j) = \sum_{i=1}^{\infty} p_{i,j}, \ j \ge 1. \end{aligned} \tag{2.2}$$

我们称 X 的分布 $\{p_j\}$, Y 的分布 $\{q_j\}$ 为 (X, Y) 的**边缘分布**.

概率分布表

- 当 (X, Y) 的概率分布的规律性不强, 或不能用 (2.2) 明确表达时, 还可以用表格的形式表达如下.

$p_{i,j}$	y_1	y_2	y_3	$\cdots$	y_n	$\cdots$	$\{p_i\}$
x_1	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$\cdots$	$p_{1,n}$	$\cdots$	p_1
x_2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$\cdots$	$p_{2,n}$	$\cdots$	p_2
x_3	$p_{3,1}$	$p_{3,2}$	$p_{3,3}$	$\cdots$	$p_{3,n}$	$\cdots$	p_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\{q_j\}$	q_1	q_2	q_3	$\cdots$	q_n	$\cdots$	1

- 这时 p_i 是其所在行中 $p_{i,j}$ 之和, q_j 是其所在列的 $p_{i,j}$ 之和.

离散型随机变量的独立性

- **定理 2.1** 设离散型随机向量 (X,Y) 的所有不同取值是

$$(x_i, y_j), \quad i,j \geq 1,$$

则 X, Y 相互独立的充分必要条件是对任何 (x_i, y_j),

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j). \tag{2.3}$$

定理 2.1 证明

- 如果 (2.3) 成立, 则对任何 $x, y \in R$, 利用概率的可列可加性得到

$$\begin{aligned} P(X \leq x, Y \leq y) =& P\Big(\bigcup\nolimits_{i:x_i \leq x} \{X = x_i\}, \bigcup\nolimits_{j:y_j \leq y} \{Y = y_j\}\Big) \\ =& \sum_{i:x_i \leq x} \sum_{j:y_j \leq y} P(X = x_i, Y = y_j) \\ =& \sum_{i:x_i \leq x} \sum_{j:y_j \leq y} P(X = x_i)P(Y = y_j) \\ =& \sum_{i:x_i \leq x} P(X = x_i) \sum_{j:y_j \leq y} P(Y = y_j) \\ =& P(X \leq x)P(Y \leq y). \end{aligned}$$

于是 X,Y 独立.

- 现在设 X, Y 独立. 对于 $A = \{x_k | k = i\}$ 和 $B = \{y_k | k = j\}$ 由定理 1.1 知道 $\{X = x_i\} = \{X \in A\}$ 与 $\{Y = y_j\} = \{Y \in B\}$ 独立, 所以有 (2.3) 成立.

离散型随机向量

- 如果 $X_1, X_2, \cdots, X_n$ 都是离散型的随机变量, 就称 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 是离散型随机向量. 如果 $\boldsymbol{X}$ 所有的不同取值是

$$(x_{1j_1}, x_{2j_2}, \cdots, x_{nj_n}),\ j_1, j_2, \cdots, j_n \geq 1,$$

就称

$$\begin{aligned} & p(j_1, j_2, \cdots, j_n) \\ =& P\big(X_1 = x_{1j_1}, \cdots, X_n = x_{nj_n}\big),\ j_1, j_2, \cdots, j_n \geq 1, \end{aligned}$$

是 $\boldsymbol{X}$ 的联合概率分布.

多项分布

- **例 2.1(多项分布)** 设 $A_1, A_2, \ldots, A_r$ 是试验 S 的完备事件组，$p_i = P(A_i)$。对试验 S 进行 n 次独立重复试验时，用 X_i 表示结果 A_i 发生的次数。
- 令随机向量 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$，则 $X_1 + X_2 + \cdots + X_r = n$，这 r 个随机变量可以用 $(X_1, X_2, \ldots, X_{r-1})$ 表示。
- $\boldsymbol{X}$ 的概率分布为

$$\begin{aligned} &P(X_1 = k_1, X_2 = k_2, \ldots, X_r = k_r) \\ =&\frac{n!}{k_1!k_2!\ldots k_r!} p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r}, \end{aligned} \tag{2.4}$$

  其中 k_i 在 $\{0, 1, \ldots, n\}$ 内取值且 $\sum_{i=1}^{r} k_i = n$。$\sum_{i=1}^{r} p_i = 1$。
- **证明:** 关于 n 用归纳法。
- $n = 1$ 时，$(k_1, k_2, \ldots, k_r)$ 中仅有一个等于 1，其它等于零。若 $k_i = 1$，则仅有的一次试验结果为 A_i，所以 (2.4) 左边即 $P(A_i) = p_i$，显然右边等于 p_i。
- 设 (2.4) 对 $n-1$ 成立。用 Y_i 表示在前 $n-1$ 次中 A_i 发生的次数，用 B_i 表示第 n 次试验结果为 A_i。
- 用全概率公式：

$$\begin{aligned} &P(X_1 = k_1, X_2 = k_2, \ldots, X_r = k_r) \\ =&\sum_{i=1}^{r} P(X_1 = k_1, X_2 = k_2, \ldots, X_r = k_r \,|\, B_i)\, P(B_i) \\ =&\sum_{i=1}^{r} P(Y_i = k_i - 1,\ Y_j = k_j, j \neq i \,|\, B_i)\, p_i \\ =&\sum_{i=1}^{r} \frac{(n-1)!}{k_1!\ldots k_{i-1}!(k_i-1)!k_{i+1}!\ldots k_r!} p_1^{k_1} \ldots p_{i-1}^{k_{i-1}} p_i^{k_i - 1} p_{i+1}^{k_{i+1}} \ldots p_r^{k_r} \cdot p_i \\ =&\sum_{i=1}^{r} \frac{k_i}{n} \cdot \frac{n!}{k_1!\ldots k_{i-1}!k_i!k_{i+1}!\ldots k_r!} p_1^{k_1} \ldots p_{i-1}^{k_{i-1}} p_i^{k_i} p_{i+1}^{k_{i+1}} \ldots p_r^{k_r} \\ =&\sum_{i=1}^{r} \frac{n!}{k_1!k_2!\ldots k_r!} p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r}\ . \end{aligned}$$

- 当 $r=2$ 时，(X_1,X_2) 完全由 X_1 决定，$X_1\sim \mathrm{B}(n,p_1)$。
- 当 $r>2$ 时，X_i 的边缘分布显然为 $\mathrm{B}(n,p_i)$。
- 当 $r=3$ 时，(X_1,X_2,X_3) 完全由 (X_1,X_2) 决定，(X_1,X_2) 的联合分布为

$$\begin{aligned}&P(X_1=k_1,X_2=k_2)\\=&\frac{n!}{k_1!k_2!(n-k_1-k_2)!}p_1^{k_1}p_2^{k_2}\dots(1-p_1-p_2)^{n-k_1-k_2},\\&0\le k_1,k_2\le n,\ k_1+k_2\le n,\ \ p_1+p_2<1.\end{aligned}$$

3.3 连续型随机向量及其分布

联合概率密度

- **定义 3.1** 设 (X,Y) 是随机向量, 如果有 R^2 上的非负可积函数 $f(x,y)$ 使得对 R^2 的所有长方形子集

$$D=\{\ (x,y)\mid a<x\le b, c<y\le d\ \} \tag{3.1}$$

有

$$P((X,Y)\in D)=\int\!\!\int_D f(x,y)\,dxdy, \tag{3.2}$$

就称 (X,Y) 是连续型随机向量, 并称 $f(x,y)$ 是 (X,Y) 的**联合概率密度**或**联合密度** (joint density).

- 按照上述定义, 连续型随机向量有概率密度, 没有概率密度的随机向量不是连续型随机向量.
- 注意，与一元连续型随机变量的定义类似，不是可以在 $\mathbb{R}^2$ 连续取值的二元随机向量就可以称为连续型随机向量，必须要有联合密度。
- 例如，设 $X\sim \mathrm{U}(0,1)$, $Y=1-X$，则 $Y\sim \mathrm{U}(0,1)$，但是 (X,Y) 取值都在线段 $\{(x,y)|\ 0<x<1, y=1-x\}$ 上，任何二元可积函数在此线段上的积分等于零。
- 设 $f(x,y)$ 是 (X,Y) 的概率密度. 可以证明对 $\mathbb{R}^2$ 的子区域 B, 有

$$P((X,Y)\in B)=\int\!\!\int_B f(x,y)dx\,dy. \tag{3.3}$$

于是有

$$\int\!\!\int_{R^2} f(x,y)dx\,dy=P((X,Y)\in\mathbb{R}^2)=1. \tag{3.4}$$

Fubini 定理

- **定理 3.1** (Fubini 定理) 设 D 是 R^n 的子区域, $\varphi(x_1, x_2, \cdots, x_n)$ 是 D 上的非负函数或满足

$$\int\int\cdots\int_D |\varphi(x_1, x_2, \cdots, x_n)|dx_1dx_2\cdots dx_n < \infty,$$

则对区域 D 上的 n 重积分

$$\int\int\cdots\int_D \varphi(x_1, x_2, \cdots, x_n)dx_1dx_2\cdots dx_n$$

可以进行累次积分计算, 且积分的次序可以交换.

边缘密度

- 设 $f(x, y)$ 是随机向量 (X, Y) 的概率密度, 则 X 和 Y 也都是连续型随机变量，我们称 X, Y 各自的概率密度为 $f(x, y)$ 或 (X, Y) 的边缘密度 (marginal density).

- 对任何 $a < b$, 有

$$\begin{aligned} P(a < X \le b) =& P(a < X \le b, Y < \infty) \\ =& \int_a^b \Big(\int_{-\infty}^{\infty} f(x, y)dy \Big) dx. \end{aligned}$$

由概率密度的定义知道 X 有边缘密度

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy. \tag{3.5}$$

- 完全对称地得到 Y 的边缘函数

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

例: 平面上的均匀分布

- 设 D 是 $\mathbb{R}^2$ 的子区域, 面积 $m(D) \in (0, \infty)$. 如果 (X, Y) 有密度函数

$$f(x, y) = \begin{cases} \frac{1}{m(D)}, & (x, y) \in D, \\ 0, & (x, y) \notin D, \end{cases} \tag{3.6}$$

就称 (X, Y) 服从 D 上的均匀分布, 记做 $(X, Y) \sim \mathrm{U}(D)$.

- **例 3.1** 设 (X,Y) 在单位圆 $D=\{(x,y)|x^2+y^2\le 1\}$ 内均匀分布, 求 X 和 Y 的概率密度.

- **解** 用 I_D 表示 D 的示性函数, 则 (X,Y) 有联合密度 $f(x,y)=(1/\pi)\mathrm{I}_D$. X 只在 $[-1,1]$ 中取值. 由 (3.5) 知道

$$\begin{aligned} f_X(x) =& \int_{-\infty}^{\infty} f(x,y)dy \\ =& \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{I}_{\{x^2+y^2\le 1\}}dy \\ =& \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{I}_{\{\,|y|\le\sqrt{1-x^2}\,\}}dy \\ =& \frac{2}{\pi}\sqrt{1-x^2},\ |x|\le 1. \end{aligned}$$

- 同理得到 Y 的密度函数

$$f_Y(y)=\frac{2}{\pi}\sqrt{1-y^2},\ |y|\le 1.$$

联合分布与联合密度

- **定理 3.2** 设 (X,Y) 有连续的分布函数 $F(x,y)$, 定义

$$f(x,y)=\begin{cases}\frac{\partial^2 F(x,y)}{\partial x\partial y}, & \text{当该混合偏导数存在,}\\ 0, & \text{其他.}\end{cases} \tag{3.7}$$

  如果

$$\int\int_{R^2} f(x,y)dxdy=1,$$

  则 $f(x,y)$ 是 (X,Y) 的联合密度.

- 证明略.

没有密度的例子

- **例 3.2** 存在随机向量 (X,Y), 它有连续的联合分布函数, 但不是连续型随机变量.

- **解** 设 X 在 $[0,1]$ 上均匀分布, $Y=X$. 则 (X,Y) 有连续的联合分布

函数 (见图 3.3.1)

$$\begin{aligned} F(x,y) =& P(X \le x, X \le y) \\ =& P(X \le \min(x,y)\ ) \\ =& \begin{cases} 0, & \min(x,y) \le 0, \\ \min(x,y), & \min(x,y) \in (0,1], \\ 1, & \min(x,y) > 1. \end{cases} \end{aligned}$$

- (X,Y) 只在 $D = \{(x,y)|0 \le x = y \le 1\}$ 上取值. 如果 (X,Y) 有联合密度 $f(x,y)$, 则有矛盾的结果

$$1 = P((X,Y) \in D) = \int_D f(x,y)dxdy = 0.$$

所以 (X,Y) 没有联合密度, 从而不是连续型随机变量.

- 例 3.2 说明: $F(x,y)$ 连续, 除去有限条直线外, $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ 存在且连续的条件还不能保证 (X,Y) 有联合密度.

连续型随机变量独立性定理

- **定理 3.3** 设 X,Y 分别有概率密度 $f_X(x)$, $f_Y(y)$. 则 X,Y 独立的充分必要条件是随机向量 (X,Y) 有联合密度

$$f(x,y) = f_X(x)f_Y(y). \tag{3.8}$$

- **证明** 如果 (3.8) 是 (X,Y) 的联合密度, 则有

$$\begin{aligned} P(X \le x, Y \le y) =& \int_{-\infty}^{x} \Big(\int_{-\infty}^{y} f_X(s)f_Y(t)dt \Big) ds \\ =& \int_{-\infty}^{x} f_X(s)ds \int_{-\infty}^{y} f_Y(t)dt \\ =& P(X \le x)P(Y \le y). \end{aligned}$$

由定义 1.1 知道 X,Y 独立.

- 如果 X,Y 独立, 对 $a \le b$, $c \le d$, 利用 Fubini 定理得到

$$\begin{aligned} & P(a < X \le b, c < Y \le d) \\ =& P(a < X \le b)(c < Y \le d) \\ =& \int_a^b f_X(x)dx \int_c^d f_Y(y)dy \\ =& \int_a^b \int_c^d f_X(x)f_Y(y)dxdy. \end{aligned}$$

按联合分布密度定义可知 $f_X(x)f_Y(y)$ 是 (X,Y) 的联合分布密度。

均匀分布

- **例 3.3** 设 (X,Y) 在矩形 $D=\{(x,y)|a<x\le b, c<y\le d\}$ 上均匀分布. 容易计算出 X 和 Y 的概率密度如下,

$$f_X(x)=\int_{-\infty}^{\infty}\frac{\mathrm{I}_D}{m(D)}dy=\frac{1}{(b-a)}\mathrm{I}_{(a,b]},$$
$$f_Y(y)=\int_{-\infty}^{\infty}\frac{\mathrm{I}_D}{m(D)}dx=\frac{1}{(d-c)}\mathrm{I}_{(c,d]}.$$

  于是 $X\sim \mathrm{U}(a,b)$, $Y\sim \mathrm{U}(c,d)$.

- 由于

$$f_X(x)f_Y(y)=\frac{1}{m(D)}\mathrm{I}_D$$

  是 (X,Y) 的联合密度, 所以 X,Y 相互独立.

- 反之, 若 $X\sim \mathrm{U}(a,b)$, $Y\sim \mathrm{U}(c,d)$, 且 X,Y 相互独立, 则 (X,Y) 在 D 上均匀分布. 因为这时 (X,Y) 的联合密度在 D 上是常数.

- 对于非长方形上的均匀分布其分量非均匀分布。

联合概率计算的例子

- **例 3.4** 两人某天在 1 点至 2 点间独立地随机到达某地会面, 先到者等候 20 分钟后离去. 求这两人能相遇的概率.

- **解** 认为每个人在 0 至 60 分钟内等可能到达, 用 X, Y 分别表示他们的到达时间. 则 $X\sim \mathrm{U}(0,60)$, $Y\sim \mathrm{U}(0,60)$, X,Y 独立. 利用

$$f_X(x)=f_Y(x)=\begin{cases}\frac{1}{60}, & x\in(0,60),\\ 0, & x\notin(0,60),\end{cases}$$

  得到 (X,Y) 的联合密度

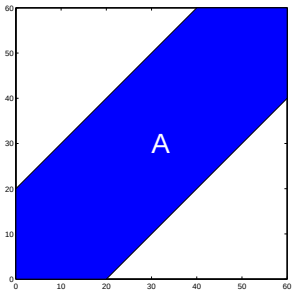
$$f(x,y)=f_X(x)f_Y(y)=\begin{cases}1/60^2, & (x,y)\in D,\\ 0, & (x,y)\notin D.\end{cases}$$

  其中 $D=\{(x,y)|0\le x,y\le 60\}$. 定义 (见图 3.3.2)

$$A=\{\ (x,y)\ \big|\ |x-y|\le 20, (x,y)\in D\ \}.$$

- 要计算的概率是

$$\begin{aligned} P(|X-Y|\le 20) =& \int\int_A f(x,y)dxdy \\ =& \frac{60^2-40^2}{60^2} = \frac{5}{9}. \end{aligned}$$



二维正态分布

- **例 3.5** (二维正态分布) 设 μ_1,μ_2 是常数, σ_1,σ_2 是正常数, ρ 是 $(-1,1)$ 中的常数.
- 如果随机向量 (X,Y) 有概率密度

$$\begin{aligned} f(x,y) =& \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2(1-\rho^2)}\Big[\frac{(x-\mu_1)^2}{\sigma_1^2} \\ & -\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\Big]\Big\}, \end{aligned} \tag{3.9}$$

  就称 (X,Y) 服从二维正态分布, 记做 $(X,Y)\sim \mathrm{N}(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$.

二维正态分布 (续)

- **例 3.6** (接例 3.5) 设 (X,Y) 有联合密度 (3.9). 证明

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

(2) X,Y 独立的充分必要条件是 $\rho = 0$.

- **解** (1) 引入

$$u = \frac{x-\mu_1}{\sigma_1},\ v = \frac{y-\mu_2}{\sigma_2},\ \mu = \rho u, \sigma = \sqrt{1-\rho^2},$$

- 则有

$$\begin{aligned}
&f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \\
=&\frac{\sigma_2}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\Big[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\Big]dv \\
=&\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\Big[-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\Big]\ \exp\Big(-\frac{u^2}{2}\Big)dv \\
=&\frac{1}{\sqrt{2\pi}\sigma_1} \exp\Big(-\frac{u^2}{2}\Big) \times \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\Big[-\frac{(v-\mu)^2}{2\sigma^2}\Big]dv \\
=&\frac{1}{\sqrt{2\pi}\sigma_1} \exp\Big(-\frac{u^2}{2}\Big) \quad \text{(注意密度积分为 1)} \\
=&\frac{1}{\sqrt{2\pi}\sigma_1} \exp\Big(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\Big).
\end{aligned}$$

于是 $X \sim N(\mu_1, \sigma_1^2)$.

- 完全对称地得到 $Y \sim N(\mu_2, \sigma_2^2)$.

- X,Y 独立的充分必要条件是

$$f_X(x) f_Y(y) = f(x,y). \tag{3.10}$$

- 当 $\rho = 0$, (3.10) 成立, 于是 X,Y 独立.

- 当 X,Y 独立, 有 (3.10) 成立, 取 $(x,y) = (\mu_1, \mu_2)$ 得到

$$\frac{1}{\sqrt{2\pi}\sigma_1}\frac{1}{\sqrt{2\pi}\sigma_2} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

于是有 $\rho = 0$.

- 即证明了 X,Y 独立的充分必要条件是 $\rho = 0$.

- 在军事武器的鉴定工作中, 如果用 (X,Y) 表示某型号导弹的弹落点, 则 (X,Y) 服从二维正态分布. 这时 (μ_1, μ_2) 是目标的地理坐标, $R = \sqrt{(X-\mu_1)^2 + (Y-\mu_2)^2}$ 是弹落点至目标的距离, 被称为**脱靶量**. σ_1^2, σ_2^2 的大小描述的是制导精度.

n 元连续型随机向量

- **定义 3.2** 设 $\boldsymbol{X}=\{X_1,X_2,\cdots,X_n\}$ 是随机向量, 如果有 R^n 上的非负可积函数 $f(x_1,x_2,\cdots,x_n)$ 使得对 R^n 的任何子立方体

$$D=\{\ (x_1,x_2,\cdots,x_n)\mid a_i<x_i\le b_i,\ 1\le i\le n\ \}$$

有

$$P(\boldsymbol{X}\in D)=\int\cdots\int_D f(x_1,x_2,\cdots,x_n)dx_1\,dx_2\cdots dx_n,$$

就称 $\boldsymbol{X}$ 是连续型随机向量, 并称 $f(x_1,x_2,\cdots,x_n)$ 是 $\boldsymbol{X}$ 的**联合概率密度函数**, 简称为**联合密度**或**概率密度**.

- 按照上述定义, 连续型随机向量有概率密度, 没有概率密度的随机向量不是连续型随机向量.

- 设 $f(x_1,x_2,\cdots,x_n)$ 是 $\boldsymbol{X}$ 的概率密度. 可以证明对 R^n 的子集 B, 有

$$P(\boldsymbol{X}\in B)=\int\cdots\int_B f(x_1,x_2,\cdots,x_n)dx_1\,dx_2\cdots dx_n.$$

于是有

$$\int\cdots\int_{R^n} f(x_1,x_2,\cdots,x_n)dx_1\,dx_2\cdots dx_n=P(\boldsymbol{X}\in R^n)=1.$$

n 个随机变量的独立性定理

- **定理 3.4** 设对每个 $i(1\le i\le n)$, 随机变量 X_i 有概率密度 $f_i(x_i)$. 则 $X_1,X_2,\cdots,X_n$ 相互独立的充分必要条件是随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 有联合密度

$$f_1(x_1)f_2(x_2)\cdots f_n(x_n),\ (x_1,x_2,\cdots,x_n)\in R^n. \tag{3.11}$$

- 独立的随机向量的值域一定是乘积集合 $\{(x_1,x_2,\ldots,x_n):\ x_i\in B_i,i=1,2,\ldots,n\}$。

3.4 随机向量函数的分布

离散型随机向量的函数—Poisson 分布例

- **例 4.1** (泊松分布的可加性) 设一个公交车站有 1 路, 2 路, $\cdots$, n 路汽车停靠. 早 7 点至 8 点之间乘 i 路车的乘客到达数 X_i 服从参数是 λ_i 的泊松分布. 计算 7 点至 8 点之间

(1) 乘 1 和 2 路汽车的到达人数 $Z_2 = X_1 + X_2$ 的概率分布,

(2) 到达总人数 $Z_n = X_1 + X_2 + \cdots + X_n$ 的概率分布.

- **解** Z_2 是取非负整数值的随机变量, 对于 $k = 0, 1, \cdots$, 有

$$\begin{aligned} P(Z_2 = k) =& P(X_1 + X_2 = k) \\ =& P(\bigcup_{i=0}^{k} \{X_1 = i, X_2 = k - i\}) \\ =& \sum_{i=0}^{k} P(X_1 = i) P(X_2 = k - i) \\ =& \sum_{i=0}^{k} \frac{\lambda_1^i}{i!} \frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_1 - \lambda_2} \qquad (X, Y\text{独立}) \\ =& \sum_{i=0}^{k} \frac{1}{k!} C_k^i \lambda_1^i \lambda_2^{k-i} e^{-\lambda_1 - \lambda_2} \\ =& \frac{(\lambda_1 + \lambda_2)^k}{k!} e^{-(\lambda_1 + \lambda_2)}. \end{aligned}$$

  说明到达的总人数 $Z_2 = X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

- 现在假设 $Z_{n-1} \sim \text{Poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_{n-1})$. 利用 Z_{n-1} 和 X_n 独立, $Z_n = Z_{n-1} + X_n$ 和 (1) 中的结果得到

$$Z_n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_n).$$

- 从例 4.1 可以得到如下的结果: 如果 $X_1, X_2, \cdots, X_n$ 相互独立, $X_i \sim \text{Poisson}(\lambda_i)$, 则 $Z_n = X_1 + X_2 + \cdots + X_n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$.

离散型随机向量—二项分布例

- **例 4.2** 设全班有 k 个同学, 在相同的条件下每个同学重复进行同一试验. 如果第 i 个同学作了 m_i 次试验, 其中试验成功的次数是 X_i. 计算全班同学的试验成功总次数

$$Z_n = X_1 + X_2 + \cdots + X_n$$

  的概率分布.

- **解** 设每次试验成功的概率是 p, 全班同学一共进行了 $m = m_1 + m_2 + \cdots + m_n$ 次独立重复试验, 试验成功的总次数 Z_n 服从二项分布 $B(m, p)$.

- 例 4.2 说明: 如果 X_i 服从二项分布 $B(m_i, p)$ 分布, $X_1, X_2, \cdots, X_n$ 相互独立, 则它们的和 $Z_n = X_1 + X_2 + \cdots + X_n$ 服从二项分布 $B(m_1 + m_2 + \cdots + m_n, p)$.
- 当然也可以按照例 4.1 的方法推导出上述结果, 但是从问题的背景出发得到的结果更加直接和容易理解.

连续型随机向量函数的分布

- 设随机向量 (X, Y) 有联合密度 $f(x, y)$, $u = u(x, y)$, $v = v(x, y)$ 是二元函数, 则
$$U = u(X, Y), V = v(X, Y)$$
是随机变量, 于是可以计算 U, V 的概率分布.

连续型随机向量函数的分布—例 4.3

- **例 4.3** 设 (X, Y) 有联合密度 $f(x, y)$, 则 $Z = X + Y$ 有密度函数
$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x)dx. \tag{4.1}$$
- 特别当 X,Y 独立时, $Z = X + Y$ 有密度函数
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)dx. \tag{4.2}$$
- **解** 对任何 $a < b$,
$$\begin{aligned} P(a < Z \le b) =& P(a < X + Y \le b) \\ =& \int_{R^2} \mathrm{I}_{\{a<x+y\le b\}} f(x, y)dxdy \\ =& \int_{-\infty}^{\infty} \Big(\int_{a-x}^{b-x} f(x, y)dy \Big) dx \\ =& \int_{-\infty}^{\infty} \Big(\int_{a}^{b} f(x, z - x)dz \Big) dx \qquad (\text{取 } y = z - x) \\ =& \int_{a}^{b} \Big(\int_{-\infty}^{\infty} f(x, z - x)dx \Big) dz. \end{aligned} \tag{4.3}$$
由密度函数的定义知道最后一式括号中以 z 为自变量的函数是 $Z = X + Y$ 的密度.
- 由于 $Z = Y + X$, 所以 Z 的密度也可以表示成
$$f_Z(z) = \int_{-\infty}^{\infty} f(z - y, y)dy. \tag{4.4}$$

连续型随机向量函数的分布—例 4.4

- **例 4.4** 设 (X,Y) 有联合密度 $f(x,y)$, 则 $V=X-Y$ 有密度函数

$$f_V(v)=\int_{-\infty}^{\infty} f(x,x-v)dx. \tag{4.5}$$

- 特别当 X,Y 独立时, $V=X-Y$ 有密度函数

$$f_V(v)=\int_{-\infty}^{\infty} f_X(x)f_Y(x-v)dx. \tag{4.6}$$

- 证明同例 4.3。

例 4.5 Reileigh 分布

- **例 4.5** 设 X, Y 独立, 都服从标准正态分布 $N(0,1)$, 求脱靶量 $Z=\sqrt{X^2+Y^2}$ 的分布.

- **解** (X,Y) 有联合密度

$$f(x,y)=\frac{1}{2\pi}\exp(-\frac{x^2+y^2}{2}). \tag{4.7}$$

  Z 在 $[0,\infty)$ 取值.

- 对 $z\geq 0$, 利用公式 (3.3) 得到 Z 的分布函数

$$\begin{aligned}
F_Z(z)=&P(\sqrt{X^2+Y^2}\leq z)\\
=&\int\int_{\{\sqrt{x^2+y^2}\leq z\}}\frac{1}{2\pi}\exp(-\frac{x^2+y^2}{2})dxdy\\
=&\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^z e^{-r^2/2}r\,dr\\
&\qquad(\text{取 } x=r\cos\theta, y=r\sin\theta, \text{ Jacobi}=r)\\
=&\int_0^z e^{-r^2/2}r\,dr.
\end{aligned}$$

- $F_Z(z)$ 连续, 求导得到 Z 的密度函数

$$f_Z(z)=ze^{-z^2/2},\ z\geq 0. \tag{4.8}$$

- (4.8) 称为瑞利 (Rayleigh) 分布密度, 于是脱靶量 Z 服从瑞利分布.

随机向量函数的联合密度定理

- **定理 4.1** 设 (X,Y) 有联合密度 $f(x,y)$, $U=u(X,Y)$, $V=v(X,Y)$ 是 (X,Y) 的函数, D 是平面上的区域使得

$$P((U,V)\in D)=1.$$

- 如果有 D 上的函数

$$\begin{cases} x_i = x_i(u,v), \\ y_i = y_i(u,v), \end{cases} \quad i=1,2,\cdots,n,$$

使得以下条件成立,

- (1) 对 $(u,v)\in D$, 有 $\{U=u,V=v\}=\bigcup_{i=1}^{n}\{X=x_i,Y=y_i\}$,

- (2) 每个

$$\begin{cases} x_i = x_i(u,v), \\ y_i = y_i(u,v) \end{cases} \tag{4.9}$$

是 D 到其值域 D_i 的可逆映射, 有连续的偏导数, 并且雅可比 (Jacobi) 行列式

$$\frac{\partial(x_i,y_i)}{\partial(u,v)}\neq 0,\ (u,v)\in D,\ i=1,2,\cdots n,$$

- (3) 集合 $D_1,D_2,\cdots,D_n$ 互不相交,

- 则 (U,V) 有联合密度.

$$g(u,v)=\begin{cases} \sum_{i=1}^{n} f(x_i(u,v),y_i(u,v))\left|\frac{\partial(x_i,y_i)}{\partial(u,v)}\right|, & (u,v)\in D, \\ 0, & (u,v)\in \overline{D}. \end{cases} \tag{4.10}$$

独立同分布

- **定义 4.1(独立同分布随机变量列)** 如果随机变量 $X_1,X_2,\cdots,X_n$ 相互独立并且有相同的分布, 就称它们**独立同分布** (independent and identically distributed), 简称为 i.i.d..

例 4.6

- **例 4.6** 设 X, Y 独立都服从标准正态分布 $N(0,1)$, (R,Θ) 由极坐标变换
$$\begin{cases} X = R\cos\Theta, \\ Y = R\sin\Theta \end{cases}$$
决定. 求 (R,Θ) 的联合密度.

- **解** (X,Y) 有联合密度 (4.7). 定义
$$D = \{(r,\theta) | r > 0, \theta \in [0, 2\pi)\}.$$
则 $P((R,\Theta) \in D) = 1$.

- $$\Delta: \quad \begin{cases} x = r\cos\theta, \\ y = r\sin\theta \end{cases}$$
建立了 D 到 $D_1 = \{(x,y) \,|\, (x,y) \neq 0\}$ 的可逆变换.

- 对 $(r,\theta) \in D$, 利用
$$\{R = r, \Theta = \theta\} = \{X = x, Y = y\}, \quad \frac{\partial(x,y)}{\partial(r,\theta)} = r > 0,$$

- 得到 (R,Θ) 的联合密度
$$\begin{aligned} g(r,\theta) &= f(r\cos\theta, r\sin\theta)r \\ &= \frac{1}{2\pi} r\exp(-r^2/2), \qquad (r,\theta) \in D. \end{aligned}$$

- R 和 Θ 分别有概率密度
$$\begin{aligned} g_R(r) &= \int_0^{2\pi} g(r,\theta)d\theta = r\exp(-\frac{r^2}{2})\mathrm{I}_{(0,\infty)}, \\ g_\Theta(\theta) &= \int_0^{\infty} g(r,\theta)dr = \frac{1}{2\pi}\mathrm{I}_{[0,2\pi)}. \end{aligned} \tag{4.11}$$

- 从
$$g(r,\theta) = g_R(r)g_\Theta(\theta), \qquad (r,\theta) \in D.$$
知道 R 和 Θ 独立. R 服从瑞利分布, Θ 服从 $[0,2\pi)$ 上的均匀分布.

例 4.7

- 设 X, Y 独立同分布, 都服从标准正态分布, 求 $U = X/Y, V = X^2+Y^2$ 的联合密度.
- **解** 设 $D = \{(u, v)|v > 0, -\infty < u < \infty\}$, 则 $P((U, V) \in D) = 1$.
- 对 $(u, v) \in D$, 函数 $x = u\sqrt{\frac{v}{1+u^2}}$, $y = \sqrt{\frac{v}{1+u^2}}$ 使得

$$\begin{aligned}\{U = u, V = v\} =&\{X/Y = u, X^2 + Y^2 = v\}\\ =&\{X = x, Y = y\} \ + \ \{X = -x, Y = -y\}.\end{aligned}$$

- 可以计算

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \sqrt{\frac{v}{1+u^2}} - \frac{u^2\sqrt{v}}{(1+u^2)^{3/2}} & \frac{u}{2\sqrt{v}\sqrt{1+u^2}} \\ -\frac{u\sqrt{v}}{(1+u^2)^{3/2}} & \frac{1}{2\sqrt{v}\sqrt{1+u^2}} \end{vmatrix}$$

$$\begin{aligned}=&\frac{1}{2(1 + u^2)} - \frac{u^2}{2(1 + u^2)^2} + \frac{u^2}{2(1 + u^2)^2}\\ =&\frac{1}{2(1 + u^2)}.\end{aligned}$$

- 再利用定理 4.1 得到 (U, V) 的联合密度

$$\begin{aligned}g(u, v) =&f(x, y)|J| + f(-x, -y)|J|\\ =&\frac{1}{2}\exp(-v/2)\frac{1}{\pi(1 + u^2)}, \quad v \geq 0, u \in (-\infty, \infty).\end{aligned}$$

- $g(u, v)$ 说明 U, V 独立, V 服从参数是 $1/2$ 的指数分布 (同时也是 $\Gamma(\frac{2}{2}, \frac{1}{2})$ 和 $\chi^2(2)$ 分布), U 有密度 $1/[\pi(1 + u^2)]$.
- 这时称 U 服从**柯西分布**.

3.5 极大极小值的分布

次序统计量

- 在研究产品的使用寿命时, 经常需要进行寿命试验.
- 用

$$X_1, X_2, \cdots, X_n \tag{5.1}$$

分别表示第 $1, 2, \cdots, n$ 件产品的使用寿命.

- 在时间 $t=0$ 时对这 n 件产品开始进行寿命试验.
- 用 $X_{(1)}$ 表示第一个寿终的产品的使用寿命, 用 $X_{(2)}$ 表示第二个寿终的产品的使用寿命, $\cdots\cdots$, 用 $X_{(n)}$ 表示第 n 个 (最后一个) 寿终的产品的使用寿命。
- 则每个 $X_{(i)}$ 都是随机变量, 并且

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}. \tag{5.2}$$

- 我们称 (5.2) 中的随机变量为 $X_1, X_2, \cdots, X_n$ 的 (从小到大) **次序统计量** (order statistics).
- 我们称 $X_{(1)}$ 是 $X_1, X_2, \cdots X_n$ 的极小值, 它是寿命最短的产品的使用寿命;
- 称 $X_{(n)}$ 是 $X_1, X_2, \cdots X_n$ 的极大值, 它是寿命最长的产品的使用寿命.

$$\begin{cases} X_{(n)} = \max(X_1, X_2, \cdots, X_n), \\ X_{(1)} = \min(X_1, X_2, \cdots, X_n). \end{cases} \tag{5.3}$$

极小值和极大值的意义

- 设华北地区第 j 年的降雨量是 X_j, 则

$$X_{(50)} = \max(X_1, X_2, \cdots, X_{50})$$

是 50 年一遇的最大降雨量, 也是 50 年中涝灾最严重的那一年的降雨量.

- 同理,

$$X_{(1)} = \min(X_1, X_2, \cdots, X_{50})$$

是 50 年中最干旱的那一年的降雨量.

- 明显, $X_{(1)} \le X_1 \le X_{(50)}$, $X_{(1)}$ 的分布和 X_1 的分布是不同的, $X_{(50)}$ 的分布和 X_1 的分布也是不同的.
- 研究随机变量 $X_{(1)}$, $X_{(50)}$ 的概率分布是有意义的.

例 5.1

- 设某地区的年降水量 $X_1, X_2, \cdots$, 是独立同分布的, 有公共的分布函数 $F(x)$ 和分段连续的密度函数 $f(x)$, 求 50 年一遇的最大, 最小降雨量的分布函数和分布密度.
- **解** 取 $n = 50$. $X_{(n)}$ 有分布函数

$$\begin{aligned} F_{\max}(x) =& P(X_{(n)} \le x) \\ =& P(X_1 \le x, X_2 \le x, \cdots, X_n \le x) \\ =& P(X_1 \le x)P(X_2 \le x)\cdots P(X_n \le x) \\ =& [F(x)]^n. \end{aligned}$$

- $F_{\max}(x)$ 连续, 在有导数的点求导数得到 X_n 的密度

$$f_{\max}(x) = F'_{\max}(x) = n[F(x)]^{n-1} f(x). \tag{5.4}$$

- $X_{(1)}$ 有分布函数

$$\begin{aligned} F_{\min}(x) =& P(X_{(1)} \le x) \\ =& 1 - P(X_{(1)} > x) \\ =& 1 - P(X_1 > x, X_2 > x, \cdots, X_n > x) \\ =& 1 - P(X_1 > x)P(X_2 > x)\cdots P(X_n > x) \\ =& 1 - [1 - F(x)]^n. \end{aligned}$$

- $F_{\min}(x)$ 连续, 在有导数的点求导数得到 $X_{(1)}$ 的密度

$$f_{\min}(x) = F'_{\min}(x) = n[1 - F(x)]^{n-1} f(x). \tag{5.5}$$

- 思考：设某地天气极端，一年可能滴雨不下，也可能仅下几场雨。这时 X_i 的分布不再是连续型。

例 5.2

- 某家庭原来有 4 个灯泡用于室内照明, 新装修后有 24 个灯泡用于室内照明. 装修入住后主人总认为灯泡更容易坏了, 试解释其中的原因.
- **解** 设所有灯泡的使用寿命相互独立, 且服从指数分布 $\mathrm{Exp}(\lambda)$. 用 X_i 表示第 i 只灯泡的使用寿命.

- 装修前等待第一个灯泡烧坏的时间长度 X 为

$$X = \min\{X_1, X_2, \cdots, X_4\}.$$

- 装修后等待第一个灯泡烧坏的时间长度 Y 为

$$Y = \min\{X_1, X_2, \cdots, X_{24}\}.$$

- 利用 $P(X > t) = e^{-4\lambda t}$, $P(Y > t) = e^{-24\lambda t}$ 可以分别得到 X 和 Y 的密度函数

$$f_X(t) = 4\lambda e^{-4\lambda t}, \text{ 和 } f_Y(t) = 24\lambda e^{-24\lambda t}.$$

- 所以 $X \sim \text{Exp}(4\lambda)$, $Y \sim \text{Exp}(24\lambda)$.

- 容易计算当 $\lambda = 1/(1500\text{小时})$ 时,

$$\begin{aligned} P(X > 400) &= 0.3442, & P(Y > 400) &= 0.0017 \\ P(X > 200) &= 0.5866, & P(Y > 200) &= 0.0408 \\ P(X > 100) &= 0.7651, & P(Y > 100) &= 0.2019. \end{aligned}$$

- 从中不难看出, Y 要比 X 随机地小很多.

- 装修前使用 200 小时不换灯泡的概率是 58.7%, 装修后使用 200 个小时不换灯泡是不大可能的.

3.6 条件分布和条件密度

条件分布和条件密度

- 设 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$, $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_m)$ 是随机向量, 本节讨论已知 $\boldsymbol{X} = (x_1, x_2, \cdots, x_m)$ 的条件下, $\boldsymbol{Y}$ 的概率分布.

- 为了叙述的简单, 我们只对 $n = m = 1$ 的情况详细讨论.

离散型随机变量的条件分布

- 设 (X, Y) 是离散型随机向量, 有概率分布

$$p_{ij} = P(X = x_i, Y = y_j) > 0, \quad i, j = 1, 2, \cdots, \tag{6.1}$$

- X, Y 分别有边缘分布

$$p_i = P(X = x_i),\ q_j = P(Y = y_j),\ i, j = 1, 2, \cdots. \tag{6.2}$$

- 对每个固定的 i, 由条件概率公式得到条件概率

$$P(Y=y_j|X=x_i)=\frac{P(X=x_i,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_i},\ j=1,2,\dots \quad (6.3)$$

- **定义 6.1** 称 (6.3) 为条件 $X=x_i$ 下, Y 的**条件概率分布**, 简称为条件分布 (conditional distribution).

例 6.1

- 甲向一个目标射击, 用 S_n 表示第 n 次击中目标时的射击次数.

- 如果甲每次击中目标的概率是 $p=1-q$, 则 $(X,Y)=(S_1,S_2)$ 有联合分布

$$\begin{aligned}&P(X=i,Y=j)\\ =&P(\text{失败 } i-1 \text{ 次，第 } i \text{ 次击中，再失败 } j-i-1 \text{ 次，第 } j \text{ 次击中})\\ =&p^2q^{j-2},\ j>i\geq 1.\end{aligned} \quad (6.4)$$

- 从问题的背景得到 X 的边缘分布（几何分布）

$$P(X=i)=pq^{i-1},\ i=1,2,\cdots,$$

- Y 的边缘分布是

$$\begin{aligned}P(Y=j)=&\sum_{i=1}^{j-1}P(X=i,Y=j)\\ =&\sum_{i=1}^{j-1}p^2q^{j-2}\\ =&(j-1)p^2q^{j-2},\qquad j=2,3,\cdots.\end{aligned}$$

- 于是对确定的 $j(\geq 2)$, 得到 X 的条件分布

$$P(X=i|Y=j)=\frac{p^2q^{j-2}}{(j-1)p^2q^{j-2}}=\frac{1}{j-1},\ 1\leq i<j. \quad (6.5)$$

- (6.5) 说明已知 $S_2=j$ 时, S_1 在 $\{1,2,\cdots,j-1\}$ 中的取值是等可能的.

连续型随机变量条件分布问题

- 北京夏季的高温闷热天气会造成北京电网的负荷过高.
- 用 X 表示夏季未来某天的最高气温, 用 Y 表示同一天北京电网的最大负荷, 可以认为 (X,Y) 是连续型随机向量, 有联合密度 $f(x,y)$.
- 如果已有对 X 的预测值 x, 在已知 $X=x$ 的条件下研究 Y 的概率分布是有实际意义的工作.
- 我们用

$$P(Y \le y|X=x)$$

表示已知 $X=x$ 的条件下, Y 的分布函数, 称为**条件分布函数**.

- 注意, 条件分布函数 $P(Y \le y|X=x)$ 就是最高气温为 x 的那天北京电网最大负荷的概率分布函数, 是有明确的意义的.
- 不能直接按条件概率公式理解。

连续型随机变量的条件分布推导

- 如何计算连续型随机变量的条件分布？
- 首先 X,Y 分别有边缘密度

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

- 对于充分小的正数 ε, 可以理解

$$P(Y \le y|x-\varepsilon < X \le x) \approx P(Y \le y|X=x).$$

- 另一方面, 如果 $f_X(x)$ 在 x 连续, $f_X(x)>0$, 并且 $\frac{\partial}{\partial x}F(x,y)$ 存在, 就

有

$$
\begin{aligned}
&\lim_{\varepsilon\to 0+} P(Y\le y|x-\varepsilon<X\le x)\\
=&\lim_{\varepsilon\to 0+}\frac{P(Y\le y, x-\varepsilon<X\le x)}{P(x-\varepsilon<X\le x)}\\
=&\lim_{\varepsilon\to 0+}\frac{F(x,y)-F(x-\varepsilon,y)}{F_X(x)-F_X(x-\varepsilon)}\\
=&\frac{\frac{\partial F(x,y)}{\partial x}}{F_X'(x)}\\
=&\frac{\frac{\partial}{\partial x}\int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)dtds}{f_X(x)}\\
=&\frac{\int_{-\infty}^{y} f(x,t)\,dt}{f_X(x)}\\
=&\int_{-\infty}^{y}\frac{f(x,t)}{f_X(x)}\,dt
\end{aligned}
$$

条件密度

- **定义 6.2** 设随机向量 (X,Y) 有联合密度 $f(x,y)$, X 有边缘密度 $f_X(x)$, 若在 x(确定的 x) 处 $f_X(x)>0$, 就称

$$P(Y\le y|X=x)=\int_{-\infty}^{y}\frac{f(x,t)}{f_X(x)}\,dt,\ y\in\mathbb{R} \tag{6.6}$$

为条件 $X=x$ 下, Y 的**条件分布函数** (conditional distribution function), 简称为条件分布, 记做 $F_{Y|X}(y|x)$.

- 称

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)},\ y\in\mathbb{R}, \tag{6.7}$$

为条件 $X=x$ 下, Y 的**条件概率密度**, 简称为条件密度 (conditional density).

条件密度和条件分布的关系

- 对使得 $f_X(x)>0$ 的 x,

$$(1)F_{Y|X}(y|x)=P(Y\le y|X=x)=\int_{-\infty}^{y}f_{Y|X}(t|x)\,dt,\ y\in\mathbb{R},$$

(2)如果$F_{Y|X}(y|x)$关于y连续,且除有限个点外有连续的导数,则

$$f_{Y|X}(y|x)=\begin{cases}\frac{\partial}{\partial y}F_{Y|X}(y|x), & \text{当偏导数存在,}\\ 0, & \text{当偏导数不存在,}\end{cases}$$

是$X=x$下,Y的条件密度。

例 6.2

- (接例 3.1)
- 设 (X,Y) 在单位圆 $D=\{(x,y)|x^2+y^2\le 1\}$ 内均匀分布, 则 X 和 Y 分别有概率密度:

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

- $f(x,y)=\frac{1}{\pi}\mathrm{I}_D$ 是 (X,Y) 的联合密度.
- 对 $x\in(-1,1)$, Y 有条件密度

$$f_{Y|X}(y|x)=\frac{\mathrm{I}_D}{\pi f_X(x)}=\frac{1}{2\sqrt{1-x^2}},\ |y|\le\sqrt{1-x^2}.$$

- 说明已知 $X=x$ 后, Y 在 $(-\sqrt{1-x^2},\sqrt{1-x^2})$ 上均匀分布.

例 6.3

- 设炮击的目标是 (μ_1,μ_2), 弹落点的坐标 (X,Y) 服从正态分布 $\mathrm{N}(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$.
- 已知 $X=x$ 时, 求 Y 的条件密度.
- **解** (X,Y) 的联合密度为

$$\begin{aligned}f(x,y)=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\Big\{-\frac{1}{2(1-\rho^2)}\Big[\frac{(x-\mu_1)^2}{\sigma_1^2}\\ &-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\Big]\Big\},\end{aligned}$$

- X 有边缘分布 $N(\mu_1,\sigma_1^2)$.

- $X = x$ 时, Y 的条件密度为

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} \exp\Big(- \frac{(y-\mu_x)^2}{2(1-\rho^2)\sigma_2^2}\Big), \tag{6.8}$$

  其中 $\mu_x = \mu_2 + (\rho\sigma_2/\sigma_1)(x - \mu_1)$.

- 说明已知 $X = x$ 时, $Y \sim N(\mu_x, (1-\rho^2)\sigma_2^2)$.

- 当 X 和 Y 独立时, $\rho = 0$, 于是 $f_{Y|X}(y|x) = f_Y(y)$.

布朗运动与正态分布

- 布朗运动描述浸没 (或悬浮) 在液体或气体中微小颗粒的运动, 这种现象由英国植物学家 Robert Brown 发现, 由 Einstein 于 1905 年作出解释: 微粒运动是由大量分子的连续碰撞造成的.

- 自 1918 年开始, Wiener 发表了一系列的论文对布朗运动进行数学的描述. 所以布朗运动又称为 Wiener 过程.

- 至今布朗运动已经是量子力学, 概率统计, 金融证券等研究中最重要的随机过程.

- 现在设花粉的微粒在液体表面由于受到水分子的连续碰撞而进行布朗运动, 运动起点的坐标是 (μ_1, μ_2). 用 (X, Y) 表示花粉在 t 时刻的坐标, 则 (X, Y) 服从二维正态分布.

第四章　数学期望和方差

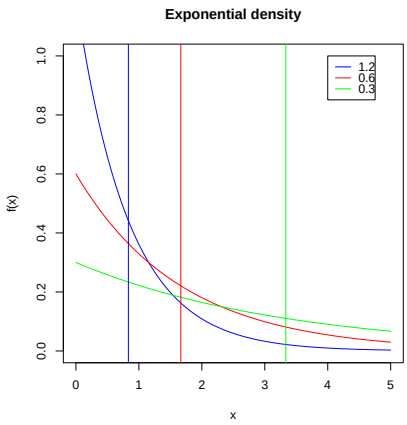
4.1　数学期望

4.1.1　数学期望概念

数学期望—引入

- 随机变量的分布函数或密度函数描述了随机变量的统计性质, 从中可以了解随机变量落入某个区间的概率, 但是还不能给人留下更直接的总体印象.
- 例如用 X 表示某计算机软件的使用寿命, 当知道 X 服从指数分布 $\mathrm{E}(\lambda)$ 后, 我们还不知道该软件的平均使用寿命是多少. 这里的平均使用寿命应当是一个实数.
- 我们需要为随机变量 X 定义一个实数, 这个数就是数学期望, 它反映随机变量的平均取值.
- 演示: 不同参数的指数分布的期望值。

不同参数的指数分布的期望值



例 1.1

- 一个班有 $n = 126$ 的学生, 期中考试后有 n_j 个同学的成绩是 j 分 $(0 \le j \le 100)$.
- 用 x_i 表示第 i 个同学的成绩, 则全班同学的平均分是

$$\mu \equiv \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{j=0}^{100} j \cdot n_j = \sum_{j=0}^{100} j\frac{n_j}{n}. \tag{1.1}$$

- 现在从班中任选一个同学, 用 X 表示这个同学的期中成绩, 则 X 有分布

$$p_j = P(X = j) = \frac{n_j}{n},\ 0 \le j \le 100. \tag{1.2}$$

- 随机变量 X 的概率分布就是该班期中考试成绩的分布, 所以 X 的数学期望应当定义成平均分 μ. 用 $\mathrm{E}(X)$ 表示 X 的数学期望时, 应当有

$$\mathrm{E}(X) = \sum_{j=0}^{100} j\frac{n_j}{n} = \sum_{j=0}^{100} jp_j.$$

例 1.2

- 设 X 有概率分布

X	1	100
p	0.01	0.99

  作为 X 的可能值的平均数, $(1 + 100)/2 = 50.5$ 并不能代表 X 的取值的平均.

- 对 X 进行 N 次独立重复观测. 由于频率是概率的估计, 所以当 N 充分大, 观测到 1 的比例大约是 0.01, 观测到 100 的比例大约是 0.99.
- 于是对 X 的单次平均观测值大约是

$$1 \times 0.01 + 100 \times 0.99 = 99.01.$$

  这就说明用

$$\mathrm{E}(X) = 1 \times 0.01 + 100 \times 0.99 = 99.01$$

  表示 X 的平均值是合理的.

数学期望定义—离散型

- **定义 1.1** 设 X 有概率分布

$$p_j = P(X = x_j),\ j = 0, 1, \cdots,$$

只要级数 $\sum_{j=0}^{\infty} |x_j| p_j$ 收敛, 就称

$$\mathrm{E}(X) = \sum_{j=0}^{\infty} x_j p_j \tag{1.3}$$

为 X 或分布 $\{p_j\}$ 的**数学期望** (expected value) 或**均值** (mean).

- 在定义 1.1 中要求 $\sum_{j=0}^{\infty} |x_j| p_j$ 收敛的原因是要使 (1.3) 中的级数有确切的意义.
- 当所有的 x_j 非负时, 如果 (1.3) 中的级数是无穷, 由 (1.3) 定义的 $\mathrm{E}(X)$ 也有明确的意义, 它表明 X 的平均取值是无穷. 这时也称 X 的数学期望是无穷.
- 不难看出, 只取有限个值的随机变量的数学期望总是存在的.

数学期望的重心解释

- 在定义 1.1 中, 将 p_j 视为横坐标 x_j 处的质量, 由

$$\sum_{j=1}^{\infty} (x_j - \mu) p_j = \sum_{j=1}^{\infty} x_j p_j - \mu = 0,$$

知道 $\{p_j\}$ 的重心是 μ.

- 所以数学期望 $\mathrm{E}(X)$ 是 X 的分布 $\{p_j\}$ 的重心.
- 对于有概率密度 $f(x)$ 的连续型随机变量 X, 我们也用 $f(x)$ 和横轴所夹面积的 (横坐标的几何) 重心定义 X 的数学期望. 设 μ 是所述的重心, 如果

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty, \tag{1.4}$$

就有

$$\int_{-\infty}^{\infty} (x - \mu) f(x) dx = \int_{-\infty}^{\infty} x f(x) dx - \mu = 0.$$

于是 $\mu = \int_{-\infty}^{\infty} x f(x) dx$ 是所述的重心.

- 参见前面不同参数的指数分布的密度图。

数学期望定义—连续型

- **定义 1.2** 设 X 是有概率密度 $f(x)$ 的随机变量, 如果下式成立,

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty, \tag{1.4}$$

就称

$$\int_{-\infty}^{\infty} x f(x) dx \tag{1.5}$$

为 X 或 $f(x)$ 的**数学期望**或**均值**.

- 和离散时的情况一样, 在定义 1.2 中要求 (1.4) 的原因是要使 (1.5) 中的积分有确切的意义.
- 当 X 非负时, 如果 (1.4) 中的积分是无穷, 由 (1.5) 定义的 $\mathrm{E}(X)$ 也有明确的意义, 它表明 X 的平均取值是无穷. 这时也称 X 的数学期望是无穷.
- 由于随机变量的数学期望由随机变量的概率分布唯一决定, 所以也可以对概率分布定义数学期望.
- 概率分布的数学期望就是以它为概率分布的随机变量的数学期望. 有相同分布的随机变量必有相同的数学期望.

一般随机变量期望

- 对于一般的随机变量 X，可以用离散随机变量期望来逼近期望。
- 取 $\varepsilon > 0$, 把数轴分为长度为 ε 的小区间，定义离散型随机变量 $X(\varepsilon)$，当 X 取值于第 i 个小区间时，$X(\varepsilon)$ 定义为小区间左端点的值：

$$X(\varepsilon) = \begin{cases} 0, & X \in [0, \varepsilon), \\ \varepsilon, & X \in [\varepsilon, 2\varepsilon), \\ \cdots, & \\ -\varepsilon, & X \in [-\varepsilon, 0), \\ -2\varepsilon, & X \in [-2\varepsilon, \varepsilon), \\ \cdots & \end{cases}$$

- 可以写成 $X(\varepsilon) = \mathrm{floor}(X/\varepsilon) \cdot \varepsilon$。
- $X(\varepsilon)$ 是离散型随机变量，$0 \le X - X(\varepsilon) < \varepsilon$。

- 设 X 是随机变量，如果 $E(X(\varepsilon))$ 存在且 $\lim_{\varepsilon\to 0+} E((X(\varepsilon))$ 存在，则定义 $EX = \lim_{\varepsilon\to 0+} E((X(\varepsilon))$，称为 X 的数学期望。
- 当 X 是离散型随机变量时，上述定义与离散型随机变量期望的公式 $EX = \sum_k x_k P(X = x_k)$ 一致。
- 当 X 是连续型随机变量，有密度 $f(x)$ 时，上述定义等价于 $EX = \int_{-\infty}^{\infty} x f(x)\, dx$。
- 参见陈家鼎、郑忠国《概率与统计》§2.6。

例 1.3

- 某省的体育彩票中, 有顺序的 7 个数字组成一个号码, 称为一注. 7 个数字中的每个数字都选自 $0, 1, 2, \cdots, 9$, 可以重复. 如果彩票一元一张, 且全体不同的彩票中只有一个大奖, 中大奖获得奖金 300 万元, 上税 20%, 甲购买一注时期望盈利多少?
- **解** 用 X 表示甲购买一注时的收益, 中大奖获利 $300 \times 80\% = 240$万, 于是

$$P(X = 240\text{万} - 1) = 10^{-7},\ P(X = -1) = 1 - 10^{-7}.$$

  X 的数学期望是

$$\mathrm{E}(X) = -1 \times (1 - 10^{-7}) + 24 \times 10^5 \times 10^{-7} = -0.76.$$

- 于是每购买一注, 甲期望获得 -0.76 元. 也就是说, 每买一注, 平均损失 0.76 元.

例 1.4

- 在澳门赌场, 有很多人在赌廿一点时顺便押对子. 其规则如下: 庄家从 6 副 (每副 52 张) 扑克中随机发给你两张. 如果你下注 a 元, 当得到的两张牌是一对时, 庄家赔你十倍, 否则输掉你的赌注. 如果你下注 100 元, 你和庄家在每局中各期望赢多少元?
- **解**: 用 X, Y 分别表示你和庄家在一局中的获利, $a = 100$. 则

$$P(X = 10a) = \frac{13C_{4\times 6}^2}{C_{52\times 6}^2} = 0.074,\quad P(X = -a) = 1 - 0.074,$$

- 于是, 你期望赢

$$\mathrm{E}X = 10a \times 0.074 - a \times (1 - 0.074) = -0.186a = -18.6(\text{元}).$$

- 庄家期望赢 18.6 元, 这是因为

$$\mathrm{E}Y = -10a \times 0.074 + a \times (1 - 0.074) = -\mathrm{E}X.$$

- 当只使用一副扑克, 可以计算出你每局期望赢 -35.32 元.

4.1.2 常见分布数学期望

两点分布 $B(1,p)$

- $P(X=1)=p$, $P(X=0)=1-p$, 则

$$\mathrm{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p.$$

- 设 A 是事件, I_A 是 A 的示性函数, 即

$$\mathrm{I}_A = \begin{cases} 1, & \text{当 } A \text{ 发生,} \\ 0, & \text{当 } A \text{ 不发生.} \end{cases}$$

则 I_A 服从两点分布, 且 $P(\mathrm{I}_A = 1) = P(A)$. 于是

$$\mathrm{E}(\mathrm{I}_A) = P(\mathrm{I}_A = 1) = P(A).$$

二项分布 $B(n,p)$

- 设 $q = 1-p$, 由

$$p_j = P(X=j) = C_n^j p^j q^{n-j},\ 0 \le j \le n,$$

- 得到

$$\begin{aligned} \mathrm{E}(X) &= \sum_{j=0}^{n} j C_n^j p^j q^{n-j} \\ &= np \sum_{j=1}^{n} C_{n-1}^{j-1} p^{j-1} q^{n-j} \qquad (\text{用 } jC_n^j = nC_{n-1}^{j-1}) \\ &= np \sum_{k=0}^{n-1} C_{n-1}^{k} p^{k} q^{n-1-k} \qquad (\text{用 } k = j-1) \\ &= np(p+q)^{n-1} \\ &= np. \end{aligned}$$

- 说明在 n 次独立重复试验中, 成功的概率 p 越大, 平均成功的次数越多.

泊松分布 Poisson(λ)

- 由
$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,\cdots,$$
- 得到
$$\mathrm{E}(X)=\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}=\lambda.$$
- 说明参数 λ 是泊松分布 Poisson(λ) 的数学期望.
- 回忆在 §2.2 的例 2.1 中, 7.5 秒内放射性钋平均放射出 3.87 个 α 粒子, 7.5 秒内放射出的粒子数 $X\sim$ Poisson(3.87).

几何分布

- 服从几何分布的随机变量 X 有离散分布
$$P(X=j)=pq^{j-1},\ j=1,2,\cdots.$$
- X 的数学期望
$$\begin{aligned}\mathrm{E}(X)=&\sum_{j=1}^{\infty}jpq^{j-1}\\ =&p\Big(\sum_{j=0}^{\infty}q^j\Big)'\\ =&p\Big(\frac{1}{1-q}\Big)'\\ =&\frac{1}{p}.\end{aligned}$$
- 结论说明单次试验中的成功概率 p 越小, 首次成功所需要的平均试验次数就越多.

指数分布 E(λ)

- 由 X 的概率密度
$$f(x)=\lambda e^{-\lambda x},\ x>0$$
- 得到
$$\mathrm{E}(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{0}^{\infty}x\lambda e^{-\lambda x}dx=\frac{1}{\lambda}.$$

- 在 §3.5 的例 5.2 中, $X_1 \sim \mathrm{E}(\lambda)$, $1/\lambda = 1500$(小时), 所以单个灯泡的平均使用寿命是 1500 小时.

- 家里用 4 只灯泡时, $X = \min\{X_1, X_2, \cdots, X_4\} \sim \mathrm{E}(4\lambda)$, 于是平均使用 $1500/4 = 375$ 小时要换一个灯泡;

- 家里用 24 只灯泡时, $Y = \min\{X_1, X_2, \cdots, X_{24}\} \sim \mathrm{E}(24\lambda)$, 平均使用 $1500/24 = 62.5$ 小时要换一个灯泡.

$\Gamma(\alpha, \lambda)$ 分布

- 由 X 的概率密度

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x > 0,$$

- 得到

$$\begin{aligned}
\mathrm{E}(X) =& \int_{-\infty}^{\infty} x f(x) dx \\
=& \int_0^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x} dx \qquad (\text{用 } t = \lambda x) \\
=& \frac{1}{\Gamma(\alpha)\lambda} \int_0^{\infty} t^\alpha e^{-t} dt \\
=& \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda} \\
=& \frac{\alpha}{\lambda}. \qquad (\text{用} \Gamma(\alpha+1) = \alpha\Gamma(\alpha))
\end{aligned}$$

- 若 X 表示寿命, 则平均寿命 $\mathrm{E}X = \alpha/\lambda$ 和 α 成正比, 和 λ 成反比.

- 这和 $f(x)$ 的形状是一致的 (演示：Gamma 分布密度。).

- 当 $\alpha = 1$, 又得到指数分布 $\mathrm{E}(\lambda)$ 的数学期望 $1/\lambda$.

对称分布的期望

- **定理 1.1** 设 X 的数学期望有限, 概率密度 $f(x)$ 关于 μ 对称: $f(\mu + x) = f(\mu - x)$, 则 $\mathrm{E}(X) = \mu$.

- **证** 这时 $g(t)=tf(t+\mu)$ 是奇函数: $g(-t)=-g(t)$. 于是

$$\begin{aligned}\mathrm{E}(X)=&\int_{-\infty}^{\infty}xf(x)dx\\=&\int_{-\infty}^{\infty}\mu f(x)dx+\int_{-\infty}^{\infty}(x-\mu)f(x-\mu+\mu)dx\\=&\mu+\int_{-\infty}^{\infty}tf(t+\mu)dt\\=&\mu.\end{aligned}$$

- 定理 1.1 的结论是自然的, 因为只要 $f(x)$ 关于 μ 对称, 则 μ 就是曲线 $f(x)$ 和 x 轴所夹面积的几何重心的横坐标.

- **推论 1.2** 正态分布 $N(\mu,\sigma^2)$ 的数学期望是 μ, 均匀分布 $\mathrm{U}(a,b)$ 的数学期望是 $(a+b)/2$.

4.2 数学期望的性质

4.2.1 随机向量函数的数学期望

随机向量函数的数学期望

- **定理 2.1** 设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 是随机向量, $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in R^n$.

(1) 如果 $\boldsymbol{X}$ 有联合密度 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$, 实函数 $g(\boldsymbol{x})$ 使得

$$\int\int\cdots\int_{R^n}|g(\boldsymbol{x})|f(\boldsymbol{x})dx_1dx_2\cdots dx_n<\infty,$$

则 $Y=g(\boldsymbol{X})$ 有数学期望

$$\mathrm{E}(Y)=\int\int\cdots\int_{R^n}g(\boldsymbol{x})f(\boldsymbol{x})dx_1dx_2\cdots dx_n;\qquad(2.1)$$

(2) 如果 $\boldsymbol{X}$ 是离散型随机向量, 有概率分布

$$p_{j_1,j_2,\cdots,j_n}=P\big(\boldsymbol{X}=(x_{j_1},x_{j_2},\cdots,x_{j_n})\big),\ j_1,j_2,\cdots,j_n\geq 1,$$

实函数 $h(\boldsymbol{x})$ 使得

$$\sum_{j_1,j_2,\cdots,j_n}|h(x_{j_1},x_{j_2},\cdots,x_{j_n})|p_{j_1,j_2,\cdots,j_n}<\infty,$$

则 $Y=h(\boldsymbol{X})$ 有数学期望

$$\mathrm{E}(Y)=\sum_{j_1,j_2,\cdots,j_n}h(x_{j_1},x_{j_2},\cdots,x_{j_n})p_{j_1,j_2,\cdots,j_n}.\qquad(2.2)$$

推论

- **推论** 设 $g(x)$ 为一元函数，若随机变量 X 有密度 $f(x)$ 且

$$\int |g(x)|f(x)dx < \infty$$

则

$$\mathrm{E}g(X) = \int_{-\infty}^{\infty} g(x)f(x)\,dx;$$

若随机变量 X 有概率分布列

$$\Pr(X = x_k) = p_k, \qquad k = 1, 2, \ldots$$

且

$$\sum_k |g(x_k)|p_k < \infty$$

则

$$\mathrm{E}g(X) = \sum_k g(x_k)p_k.$$

例 2.1

- 设 $X \sim N(0,1)$, 计算 $\mathrm{E}(X^2)$.
- **解** X 有概率密度

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2),$$

由公式 (2.1) 得到

$$\begin{aligned}
\mathrm{E}(X^2) =& \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2/2)dx \\
=& \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} x^2 \exp(-x^2/2)dx \\
=& \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} 2te^{-t} \frac{1}{\sqrt{2t}} dt \qquad (\text{取 } x = \sqrt{2t}) \\
=& \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} t^{3/2-1} e^{-t} dt = \frac{2}{\sqrt{\pi}} \Gamma(1 + \frac{1}{2}) \\
=& \frac{2}{\sqrt{\pi}} \frac{1}{2} \Gamma(\frac{1}{2}) \qquad (\text{用 } \Gamma(1+\alpha) = \alpha\Gamma(\alpha)) \\
=& 1. \qquad (\text{用 } \Gamma(1/2) = \sqrt{\pi})
\end{aligned}$$

例 2.2

- 设 X 在 $(0,\pi/2)$ 上均匀分布, 计算 $\mathrm{E}(\cos X)$.
- X 有概率密度 $f(x)=2/\pi$, $x\in(0,\pi/2)$. 用 (2.1) 得到

$$\mathrm{E}(X)=\int_{-\infty}^{\infty} f(x)\cos x\,dx=\frac{2}{\pi}\int_0^{\pi/2}\cos x\,dx=\frac{2}{\pi}.$$

例 2.3

- 设 X, Y 独立同分布且服从 $N(0,1)$, $Z=X^2+Y^2$. 计算 $\mathrm{E}(Z)$.
- **解** (X,Y) 有联合密度

$$f(x,y)=\frac{1}{2\pi}\exp\big(-\frac{x^2+y^2}{2}\big).$$

- 用公式 (2.1) 得到

$$\begin{aligned}\mathrm{E}(Z)=&\int\int_{R^2}(x^2+y^2)f(x,y)dxdy\\ =&\frac{1}{2\pi}\int\int_{R^2}(x^2+y^2)\exp\big(-\frac{x^2+y^2}{2}\big)dxdy\\ =&\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^{\infty}r^3\exp(-\frac{r^2}{2})\,dr \qquad \Big(\text{取 }\begin{cases}x=r\cos\theta,\\ y=r\sin\theta\end{cases}\Big)\\ =&2^{3/2}\int_0^{\infty}t^{3/2}e^{-t}\sqrt{\frac{1}{2t}}\,dt \qquad (\text{取 } r=\sqrt{2t})\\ =&2\int_0^{\infty}te^{-t}\,dt=2.\end{aligned}$$

例 2.4

- (X,Y) 在单位圆 $D=\{(x,y)|\,x^2+y^2\le 1\}$ 内均匀分布, 计算 $\mathrm{E}(X)$, $\mathrm{E}(Y)$.
- **解** (X,Y) 有联合密度

$$f(x,y)=\frac{1}{\pi}\mathrm{I}_D.$$

- 用公式 (2.1) 得到

$$\mathrm{E}(X)=\int\int_{R^2}xf(x,y)dxdy=\frac{1}{\pi}\int_{-1}^{1}dy\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}x\,dx=0.$$

对称地得到 $\mathrm{E}(Y)=0$.

4.2.2 数学期望的性质

数学期望的性质

- 根据定理 2.1,

$$\mathrm{E}|X| = \begin{cases} \sum_{j=1}^{\infty} |x_j| p_j, & \text{当 } p_j = P(X = x_j), \\ \int_{-\infty}^{\infty} |x| f(x)\, dx, & \text{当 } X \text{ 有密度 } f(x). \end{cases}$$

  于是, $\mathrm{E}X$ 有限的充分必要条件是 $\mathrm{E}|X| < \infty$.

- **定理 2.2** 设 $\mathrm{E}|X_j| < \infty$ $(1 \le j \le n)$, $c_0, c_1, \cdots, c_n$ 是常数, 则有以下结果.

  (1) 线性组合 $Y = c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$ 的数学期望存在, 而且

$$\mathrm{E}(Y) = c_0 + c_1 \mathrm{E}(X_1) + c_2 \mathrm{E}(X_2) + \cdots + c_n \mathrm{E}(X_n), \tag{2.3}$$

  (2) 如果 $X_1, X_2, \cdots, X_n$ 相互独立, 则 $Z = X_1 X_2 \cdots X_n$ 的数学期望存在, 并且

$$\mathrm{E}(X_1 X_2 \cdots X_n) = \mathrm{E}(X_1)\mathrm{E}(X_2) \cdots \mathrm{E}(X_n),$$

  (3) 如果 $X_1 \le X_2$, 则 $\mathrm{E}(X_1) \le \mathrm{E}(X_2)$.

定理 2.2 证明

- 只对 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 有联合密度 $f(\boldsymbol{x})$ 的情况给出证明.

- (1) 利用公式 (2.1) 得到

$$\begin{aligned} \mathrm{E}|Y| =& \int\int \cdots \int_{R^n} \Big| c_0 + \sum_{j=1}^{n} c_j x_j \Big| f(\boldsymbol{x}) dx_1 dx_2 \cdots dx_n \\ \le& |c_0| + \sum_{j=1}^{n} |c_j| \int\int \cdots \int_{R^n} |x_j| f(\boldsymbol{x}) dx_1 dx_2 \cdots dx_n \\ =& c_0 + \sum_{j=1}^{n} |c_j| \mathrm{E}|X_j| < \infty. \end{aligned}$$

- 再利用公式 (2.1) 得到

$$\begin{aligned}\mathrm{E}(Y) =& \int\int\cdots\int_{R^n}\Big(c_0+\sum_{j=1}^{n}c_jx_j\Big)f(\boldsymbol{x})dx_1dx_2\cdots dx_n\\ =&c_0+\sum_{j=1}^{n}c_j\int\int\cdots\int_{R^n}x_jf(\boldsymbol{x})dx_1dx_2\cdots dx_n\\ =&c_0+\sum_{j=1}^{n}c_j\mathrm{E}(X_j).\end{aligned}$$

- (2) 这时 $f(\boldsymbol{x})=f_1(x_1)f_2(x_2)\cdots f_n(x_n)$, 其中 $f_j(x_j)$ 是 X_j 的概率密度.

- 利用公式 (2.1) 和 Fubini 定理得到

$$\begin{aligned}\mathrm{E}|Y| =& \int\int\cdots\int_{R^n}|x_1x_2\cdots x_n|f(\boldsymbol{x})dx_1dx_2\cdots dx_n\\ =&\int_{-\infty}^{\infty}|x_1|f_1(x_1)dx_1\int_{-\infty}^{\infty}|x_2|f_2(x_2)dx_2\cdots\int_{-\infty}^{\infty}|x_n|f_n(x_n)dx_n\\ =&\mathrm{E}(|X_1|)\mathrm{E}(|X_2|)\cdots\mathrm{E}(|X_n|)<\infty.\end{aligned}$$

- 再利用公式 (2.1)) 和 Fubini 定理得到

$$\begin{aligned}\mathrm{E}(Y) =& \int\int\cdots\int_{R^n}x_1x_2\cdots x_nf(\boldsymbol{x})dx_1dx_2\cdots dx_n\\ =&\int_{-\infty}^{\infty}x_1f_1(x_1)dx_1\int_{-\infty}^{\infty}x_2f_2(x_2)dx_2\cdots\int_{-\infty}^{\infty}x_nf_n(x_n)dx_n\\ =&\mathrm{E}(X_1)\mathrm{E}(X_2)\cdots\mathrm{E}(X_n).\end{aligned}$$

- (3) 只对 $Y=X_2-X_1$ 有概率密度 $g(y)$ 的情况证明. 由 $Y\geq 0$, 知道对 $y<0$ 有 $g(y)=0$. 于是用 (1) 得到

$$\mathrm{E}(X_2)-\mathrm{E}(X_1)=\mathrm{E}(Y)=\int_0^{\infty}yg(y)\,dy\geq 0.$$

例 2.5(二项分布 $B(n,p)$)

- 设单次试验成功的概率是 p, 问 n 次独立重复试验中, 期望有几次成功?

- **解** 引入

$$X_i=\begin{cases}1, & \text{第 } i \text{ 试验成功},\\ 0, & \text{第 } i \text{ 次试验不成功}.\end{cases}$$

则 $\mathrm{E}X_i = p$, $X = (X_1 + X_2 + \cdots + X_n)$ 是 n 次试验中的成功次数, 期望的成功数是

$$\mathrm{E}X = \mathrm{E}X_1 + \mathrm{E}X_2 + \cdots + \mathrm{E}X_n = np.$$

这里 $X \sim B(n, p)$.

- 可见 X 的数学期望是指对 X 的**期望值**, 它也是 X 的平均取值.

例 2.6(超几何分布 $\mathbf{H}(n, M, N)$)

- N 件产品中有 M 件正品, 从中任取 n 件, 期望有几件正品?
- **解** 定义随机变量

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次取得正品,} \\ 0, & \text{第 } i \text{ 次取得次品.} \end{cases}$$

则无论是否有放回的抽取, 总有 $\mathrm{E}X_i = M/N$(参考 §1.6 例 6.1).

- $Y = X_1 + X_2 + \cdots + X_n$ 是抽到的正品数. 期望的正品数是

$$\mathrm{E}Y = \mathrm{E}X_1 + \mathrm{E}X_2 + \cdots + \mathrm{E}X_n = nM/N.$$

- 本例中, 有放回的抽取时,$Y \sim B(n, M/N)$.
- 无放回的抽取时, Y 服从超几何分布 $H(n, M, N)$(参考 §2.2 的 (4)).

例 2.7(信封搭配)

- 将 n 个不同的信笺随机放入 n 个写好地址的信封, 平均有几封能正确搭配?
- **解** 定义随机变量

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 封信正确搭配,} \\ 0, & \text{第 } i \text{ 封信没有正确搭配.} \end{cases}$$

则 $\mathrm{E}X_i = P(X_i = 1) = 1/n$.

- $Y = X_1 + X_2 + \cdots + X_n$ 是正确搭配的个数, 于是平均正确搭配的个数是

$$\mathrm{E}Y = \mathrm{E}X_1 + \mathrm{E}X_2 + \cdots + \mathrm{E}X_n = n/n = 1.$$

- 本例说明无论有多少个信封, 平均只有一封信能正确搭配.

例 2.8

- 如果 $\mathrm{E}|X|=0$, 则 $P(X=0)=1$.
- **证** 用 $\mathrm{I}_{\{n|X|>1\}}$ 表示事件 $\{n|X|>1\}$ 的示性函数, 利用定理 2.2 的 (3) 得到

$$\begin{aligned} P(|X|>1/n) =& P(n|X|>1) \\ =& \mathrm{E}(\mathrm{I}_{\{n|X|>1\}}) \\ \leq& \mathrm{E}(n|X|\mathrm{I}_{\{n|X|>1\}}) \\ \leq& n\mathrm{E}|X| \\ =& 0. \end{aligned}$$

- 由概率的连续性得到

$$P(|X|>0)=P(\cup_{n=1}^{\infty}\{|X|>1/n\})=\lim_{n\to\infty}P(|X|>1/n)=0.$$

  最后得到 $P(|X|=0)=1-P(|X|>0)=1$.
- 本例说明 $\mathrm{E}|X|=0$ 的充分必要条件是 $P(X=0)=1$.
- 当 $P(X=0)=1$, 我们称 $X=0$ **以概率 1 发生**, 记做 $X=0$, wp1. 这里 wp1. 表示 with probability 1.
- 完全类似地, 我们把 $P(X\leq Y)=1$ 记做 $X\leq Y$, wp1.
- 当 $P(A)=1$, 我们称 A 以概率 1 发生. 不难理解, 当 $P(A)=1$, A 在实际中必然发生.
- 以概率 1 发生又称作**几乎处处**或**几乎必然** (almost surely) 发生, 用 a.s. 表示.

例 2.9

- 设商店每销售一吨大米获利 a 元, 每库存一吨大米损失 b 元, 假设大米的销量 Y 服从指数分布 $\mathrm{E}(\lambda)$. 问库存多少吨大米才能获得最大的平均利润.
- **解** 库存量是 x 时, 利润是

$$Q(x,Y)=\begin{cases} aY-b(x-Y), & Y<x, \\ ax, & Y\geq x. \end{cases}$$

- 用 $\mathrm{I}_{\{Y<x\}}$ 表示事件 $\{Y<x\}$ 的示性函数, 用 $\mathrm{I}_{\{Y\geq x\}}$ 表示 $\{Y\geq x\}$ 的示性函数, 就可以将 $Q(x,Y)$ 写成

$$Q(x,Y)=[aY-b(x-Y)]\mathrm{I}_{\{Y<x\}}+ax\mathrm{I}_{\{Y\geq x\}}.$$

- 平均利润是

$$\begin{aligned}q(x)=&\mathrm{E}Q(x,Y)\\=&\int_{-\infty}^{\infty}Q(x,y)f_Y(y)dy\\=&\int_0^x[ay-b(x-y)]f_Y(y)dy+ax\int_x^{\infty}f_Y(y)dy\\=&\frac{a+b}{\lambda}(1-e^{-\lambda x})-bx.\qquad\text{(具体过程略)}\end{aligned}$$

- $q(x)$ 的最大值点是所要的库存数. 由

$$q'(x)=(a+b)e^{-\lambda x}-b=0,\ q''(x)=-(a+b)\lambda e^{-\lambda x}<0,$$

- 得到 $x=\lambda^{-1}\ln[(a+b)/b]$ 是 $q(x)$ 的最大值点. 于是库存 $\lambda^{-1}\ln[(a+b)/b]$ 吨大米可以获得最大平均利润.

中心矩和原点矩

- **定义 2.1** 设 X 是随机变量, m 是正整数. 如果 $\mathrm{E}|X|^m<\infty$, 就称 $\mathrm{E}X^m$ 为 X 的 m 阶**原点矩**, 称 $\mathrm{E}(X-\mathrm{E}X)^m$ 为 X 的 m 阶**中心矩**. 当 $m>2$ 时, 我们将原点矩和中心矩统称为高阶矩.

4.3 随机变量的方差

方差定义

- 方差用来描述分布的分散程度，或宽窄。
- **定义 3.1** 如果随机变量 X 的数学期望 $\mu=\mathrm{E}X$ 有限, 就称

$$\mathrm{E}(X-\mu)^2 \tag{3.1}$$

为 X 的**方差** (variance), 记做 $\mathrm{Var}(X)$ 或 σ_{XX}. 当 $\mathrm{Var}(X)<\infty$, 称 X 的方差有限. 称 $\sigma_X=\sqrt{\mathrm{Var}(X)}$ 为 X 的**标准差**.

- 设 X 是对长度为 μ 的物体的测量值, 则 $X-\mu$ 是测量误差, $(X-\mu)^2$ 是测量误差的平方.
- 如果测量仪器无系统偏差 (即 $\mathrm{E}X=\mu$), 则 $\mathrm{E}(X-\mu)^2$ 是测量误差平方的平均, 称为均方误差.

方差的计算公式

- 当 X 有离散分布 $P(X=x_j)$, $j=1,2,\cdots$ 时, 利用公式 (2.2) 得到

$$\mathrm{Var}(X)=\sum_{j=1}^{\infty}(x_j-\mu)^2P(X=x_j).$$

- 当 X 有概率密度 $f(x)$ 时, 利用公式 (2.1) 得到

$$\mathrm{Var}(X)=\int_{-\infty}^{\infty}(x-\mu)^2f(x)\,dx.$$

- 随机变量 X 的方差 $\mathrm{Var}(X)$ 由 X 的分布唯一决定.

- X 的方差描述了 X 的分散程度, $\mathrm{Var}(X)$ 越小, 说明 X 在数学期望 μ 附近越集中.

- 特别当 $\mathrm{Var}(X)=0$ 时, 由例 2.8 知道 $P(X=\mu)=1$.

- 利用数学期望的线性性质得到,

$$\mathrm{Var}(X)=\mathrm{E}(X^2-2X\mu+\mu^2)=\mathrm{E}X^2-(\mathrm{E}X)^2. \tag{3.2}$$

  (3.2) 是计算方差的常用公式.

两点分布 $\mathbf{B}(1,p)$ 的方差

- $P(X=1)=p$, $P(X=0)=1-p$.

- 由 $X^2=X$ 和 $\mathrm{E}X=p$, 得到

- $$\mathrm{Var}(X)=\mathrm{E}X^2-(\mathrm{E}X)^2=p-p^2=pq.$$

二项分布 $\mathbf{B}(n,p)$ 的方差

- 设 $q=1-p$, 由 $\mathrm{E}X=np$,

$$P(X=j)=C_n^jp^jq^{n-j},\ 0\le j\le n,$$

  和计算随机向量函数数学期望的公式 (2.2) 得到

- $$\begin{aligned}\mathrm{E}(X^2) =& \ \mathrm{E}[X(X-1)] + \mathrm{E}X\\ =&\sum_{j=0}^{n} C_n^j j(j-1)p^j q^{n-j} + np\\ =&p^2 \frac{d^2}{dx^2}\sum_{j=0}^{n} C_n^j x^j q^{n-j}\Big|_{x=p} + np\\ =&p^2 \frac{d^2}{dx^2}(x+q)^n\Big|_{x=p} + np\\ =&n(n-1)p^2 + np.\end{aligned}$$

- 最后用方差的计算公式 (3.2) 得到

$$\mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = npq.$$

- 后面讲了定理 3.1 后可以用 B(1,p) 来计算 B(n,p) 的方差。

泊松分布 Poission(λ) 的方差

- 由 $\mathrm{E}X = \lambda$,

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

和函数期望公式 (2.2) 得到

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

- 用方差的展开计算公式 (3.2) 得到

$$\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

几何分布

- 服从几何分布的随机变量 X 有离散分布

$$P(X=j) = pq^{j-1},\ j = 1, 2, \cdots.$$

- 用 $\mathrm{E}X = 1/p$ 和公式函数期望公式 (2.2) 得到

$$\begin{aligned}\mathrm{E}(X^2) =& \mathrm{E}[X(X-1)] + \mathrm{E}X \\ =& \sum_{j=1}^{\infty} j(j-1)pq^{j-1} + \frac{1}{p} \\ =& pq\Big(\sum_{j=0}^{\infty} q^j\Big)'' + \frac{1}{p} \\ =& pq\Big(\frac{1}{1-q}\Big)'' + \frac{1}{p} \\ =& \frac{2pq}{(1-q)^3} + \frac{1}{p} \\ =& \frac{2q}{p^2} + \frac{1}{p}.\end{aligned}$$

- 最后用方差的展开计算公式 (3.2) 得到

$$\mathrm{Var}(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

均匀分布 $\mathbf{U}(a,b)$

- 由 X 的概率密度

$$f(x) = \frac{1}{b-a}\mathrm{I}_{(a,b)},$$

- 数学期望 $\mathrm{E}X = (a+b)/2$, 和

- $$\mathrm{E}X^2 = \int_a^b \frac{x^2}{b-a}dx = \frac{b^3-a^3}{3(b-a)}.$$

  得到

- $$\mathrm{Var}(X) = \frac{b^3-a^3}{3(b-a)} - \Big(\frac{a+b}{2}\Big)^2 = \frac{(b-a)^2}{12}.$$

指数分布 $\mathbf{E}(\lambda)$

- X 有数学期望 $\mathrm{E}X = 1/\lambda$ 和概率密度

$$f(x) = \lambda e^{-\lambda x},\ x > 0.$$

- 于是由

$$\begin{aligned}\mathrm{E}(X^2)=&\int_0^\infty x^2\lambda e^{-\lambda x}dx\\ =&\frac{1}{\lambda^2}\int_0^\infty t^2e^{-t}dt \qquad (\text{取 } x=t/\lambda)\\ =&\frac{1}{\lambda^2}\Gamma(3)\\ =&\frac{2}{\lambda^2},\end{aligned}$$

- 得到

$$\mathrm{Var}(X)=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}.$$

正态分布 $\mathbf{N}(\mu,\sigma^2)$

- X 有数学期望 $\mathrm{E}X=\mu$
- 和概率密度

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

- 于是

$$\begin{aligned}\mathrm{Var}(X)=&\int_{-\infty}^\infty(x-\mu)^2f(x)dx\\ =&\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^\infty(x-\mu)^2\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)dx\\ =&\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^\infty t^2\exp\Big(-\frac{t^2}{2}\Big)dt \qquad (\text{取 } x-\mu=\sigma t)\\ =&\sigma^2. \qquad (\text{用例 2.1})\end{aligned}$$

- 现在我们知道了正态分布 $N(\mu,\sigma^2)$ 中的 μ 和 σ^2 就是该正态分布的数学期望和方差. 如果 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$, 则从 §3.3 的例 3.6 知道 $X\sim(\mu_1,\sigma_1^2)$,$Y\sim(\mu_2,\sigma_2^2)$, 于是得到 $\mu_1=\mathrm{E}X$, $\mu_2=\mathrm{E}Y$, $\sigma_1^2=\mathrm{Var}(X)$, $\sigma_2^2=\mathrm{Var}(Y)$.

方差的性质

- **定理 3.1** 设 a,b,c 是常数, $\mathrm{E}X=\mu$, $\mathrm{Var}(X)<\infty$, $\mu_j=\mathrm{E}X_j$, $\mathrm{Var}(X_j)<\infty(1\le j\le n)$, 则

  (1) $\mathrm{Var}(a+bX)=b^2\mathrm{Var}(X)$,

  (2) $\mathrm{Var}(X)=\mathrm{E}(X-\mu)^2<\mathrm{E}(X-c)^2$, 只要 $c\ne\mu$,

(3) $\mathrm{Var}(X)=0$ 的充分必要条件是 $P(X=\mu)=1$,

(4) $\mathrm{Var}(\sum_{j=1}^{n}X_j)=\sum_{i=1}^{n}\sum_{j=1}^{n}[\mathrm{E}(X_iX_j)-\mu_i\mu_j]$,

(5) 当 $X_1,X_2,\cdots,X_n$ 相互独立, $\mathrm{Var}(\sum_{j=1}^{n}X_j)=\sum_{j=1}^{n}\mathrm{Var}(X_j)$.

定理证明

- (1) 由方差的定义得到

$$\begin{aligned}\mathrm{Var}(a+bX)=&\mathrm{E}[(a+bX)-\mathrm{E}(a+bX)]^2\\=&\mathrm{E}[a+bX-(a+b\mu)]^2\\=&\mathrm{E}[b^2(X-\mu)^2]\\=&b^2\mathrm{Var}(X)\end{aligned}$$

- (2) 对 $c\neq\mu$, 由

$$\begin{aligned}\mathrm{E}(X-c)^2=&\mathrm{E}[(X-\mu)+(\mu-c)]^2\\=&\mathrm{E}(X-\mu)^2+2\mathrm{E}(X-\mu)(\mu-c)+\mathrm{E}(\mu-c)^2\\=&\mathrm{Var}(X)+(\mu-c)^2,\end{aligned}$$

知道 (2) 成立.

- (3) $\mathrm{Var}(X)=0$ 即 $\mathrm{E}(X-\mu)^2=0$, 由例 2.8 其充分必要条件为 $P((X-\mu)^2=0)=1$, 即 $P(X=\mu)=1$。

- (4)

$$\begin{aligned}\mathrm{Var}(\sum_j x_j)=&\mathrm{E}\left[\sum_j(X_j-\mu_j)\right]^2\\=&\sum_i\sum_j\mathrm{E}\left[(X_i-\mu_i)(X_j-\mu_j)\right]\\=&\sum_i\sum_j\left[\mathrm{E}(X_iX_j)-\mu_i\mu_j\right]\end{aligned}$$

- (5) 独立时

$$\begin{aligned}\mathrm{Var}(\sum_j X_j) =& \mathrm{E}\left[\sum_j (X_j-\mu_j)\right]^2 \\ =& \sum_j \mathrm{E}(X_j-\mu_j)^2 + 2\sum_{i<j}\mathrm{E}\left[(X_i-\mu_i)(X_j-\mu_j)\right] \\ =& \sum_j \mathrm{Var}(X_i) + 2\sum_{i<j}\left[\mathrm{E}(X_i-\mu_i)\mathrm{E}(X_j-\mu_j)\right] \\ =& \sum_j \mathrm{Var}(X_i)\end{aligned}$$

- 在性质 (1) 中取 $b=1$ 得到 $\mathrm{Var}(a+X)=\mathrm{Var}(X)$, 说明对随机变量进行常数的平移后, 随机变量的分散程度不变;
- 取 $a=0$ 得到 $\mathrm{Var}(bX)=b^2\mathrm{Var}(X)$, 说明将 X 扩大 b 倍后, 方差扩大 b^2 倍.
- (2) 说明随机变量 X 在均方误差的意义下距离 μ 最近.
- (3) 说明除了以概率 1 等于常数的随机变量外, 任何随机变量的方差都大于零.
- 以后无特殊说明, 都认为所述随机变量的方差大于零.

例 3.2

- 设 $\mathrm{Var}(X)=\sigma^2<\infty$, $Y=(X-\mathrm{E}X)/\sigma$,
- 则 $\mathrm{E}Y=0$, $\mathrm{Var}(Y)=1$.
- 这时称 Y 是 X 的标准化.
- 特别地, 当 $X\sim N(\mu,\sigma^2)$, $Y\sim N(0,1)$.

例 3.3

- 设 $X_1,X_2,\cdots,X_n$ 相互独立, 有共同的方差 $\sigma^2<\infty$, 则

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{j=1}^n X_j\Big)=\frac{1}{n}\sigma^2.$$

- 如果 X_i 是第 i 次测量重量为 μ 的物体时的测量值, 测量的均方误差是 $\mathrm{Var}(X_i)=\sigma^2$, 则用 n 次测量的平均值作为 μ 的测量值时均方误差降低 n 倍. 说明只要测量仪器没有系统偏差 (指 $\mathrm{E}X=\mu$), 测量精度总可以通过多次测量的平均来改进.

例 3.4

- (接例 2.5, 二项分布 $B(n,p)$)
- 设 $X_1, X_2, \cdots, X_n$ 独立同分布, 都服从两点分布 $B(1,p)$,
- 则 $Y = X_1 + X_2 + \cdots + X_n \sim B(n,p)$.
- 利用 $\mathrm{Var}(X_i) = pq$ 和定理 3.1 的 (5) 得到

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = npq.$$

例 3.5

- (接例 2.6, 超几何分布)
- 设 N 件产品中有 M 件正品, 无放回地从中依次取 n 件, 用 Y 表示取得的正品数, 则 $Y \sim H(n,M,N)$. 求 $\mathrm{Var}(Y)$.
- **解** 设 $X_1, X_2, \cdots, X_n$ 在例 2.6 中定义, 即 X_i 为第 i 次抽取结果的两点分布，则 $\mathrm{E}X_i = M/N$.
- 对 $i < j$, X_jX_i 服从两点分布.
- 利用抽签的原理 (参考 §1.6 的例 6.1) 知道

$$\begin{aligned}\mathrm{E}(X_jX_i) =& P(X_jX_i = 1) = P(X_j = 1, X_i = 1)\\ =& P(X_j = 1|X_i = 1)P(X_i = 1)\\ =& \frac{M-1}{N-1}\frac{M}{N}.\end{aligned}$$

- 再利用 $Y = (X_1 + X_2 + \cdots + X_n)$ 和定理 3.1 的 (4) 得到

$$\begin{aligned}\mathrm{Var}(Y) =& \sum_{j=1}^{n}\sum_{i=1}^{n}[\mathrm{E}(X_iX_j) - \mathrm{E}X_i\mathrm{E}X_j]\\ =& n\mathrm{E}X_1 + n(n-1)\mathrm{E}(X_1X_2) - n^2(\mathrm{E}X_1)^2 \qquad (\text{用 } X_1^2 = X_1)\\ =& n\frac{M}{N} + n(n-1)\frac{M-1}{N-1}\frac{M}{N} - n^2\Big(\frac{M}{N}\Big)^2\\ =& n\frac{M}{N}\Big(1-\frac{M}{N}\Big)\frac{N-n}{N-1}.\end{aligned}$$

- 特别当 $n = N$, $\mathrm{Var}(Y) = 0$. 实际上这时 $Y = \mathrm{E}Y = M$.
- 本例中如果采用有放回的抽样, 则 $Y \sim B(n, M/N)$, Y 的方差 $n\frac{M}{N}(1-\frac{M}{N})$ 要更大. 说明对均值的估计, 无放回的抽取要比有放回的抽取方差小.

Markov 不等式

- **定理 3.2** (**马尔柯夫 (Markov) 不等式**) 对随机变量 X 和 $\varepsilon > 0$, 有

$$P(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^{\alpha}} \mathrm{E}|X|^{\alpha},\ \alpha > 0. \tag{3.3}$$

- 作为定理 3.2 的直接推论, 取 $\alpha = 2$, 用 $X - \mathrm{E}X$ 代替 X 就得到

$$P(|X - \mathrm{E}X| \geq \varepsilon) \leq \frac{1}{\varepsilon^{2}} \mathrm{Var}(X),\ \varepsilon > 0. \tag{3.4}$$

- **证** 用 $\mathrm{I}[A]$ 表示事件 A 的示性函数, 对任何正数 α,

$$\begin{aligned} P(|X| \geq \varepsilon) =& \mathrm{EI}[|X| \geq \varepsilon] \\ \leq& \mathrm{E}\big(\frac{|X|^{\alpha}}{\varepsilon^{\alpha}} \mathrm{I}[|X| \geq \varepsilon]\big) \\ \leq& \mathrm{E}\frac{|X|^{\alpha}}{\varepsilon^{\alpha}} = \frac{1}{\varepsilon^{\alpha}} \mathrm{E}|X|^{\alpha}. \end{aligned}$$

- 人们称 $\alpha = 2$ 时的马尔柯夫不等式为**切比雪夫** (Chebyshev) 不等式.

内积不等式

- **定理 3.3** (**内积不等式**) 设 $\mathrm{E}X^2 < \infty$, $\mathrm{E}Y^2 < \infty$, 则有

$$|\mathrm{E}(XY)| \leq \sqrt{\mathrm{E}X^2\, \mathrm{E}Y^2}. \tag{3.5}$$

  (3.5) 中等号成立的充分必要条件是有不全为零的常数 a, b, 使得 $P(aX + bY = 0) = 1$.

- **证** 对于不全为零的常数 a, b, 二次型

$$\begin{aligned} \mathrm{E}(aX + bY)^2 =& a^2 \mathrm{E}X^2 + 2ab\mathrm{E}(XY) + b^2 \mathrm{E}Y^2 \\ =& (a, b)\Sigma(a, b)^{\mathrm{T}} \geq 0. \end{aligned} \tag{3.6}$$

  其中

$$\Sigma = \begin{pmatrix} \mathrm{E}X^2 & \mathrm{E}(XY) \\ \mathrm{E}(XY) & \mathrm{E}Y^2 \end{pmatrix}.$$

  (3.6) 说明矩阵 Σ 非负定，于是 $|\Sigma| \geq 0$。从 $\det(\Sigma) = \mathrm{E}X^2\, \mathrm{E}Y^2 - [\mathrm{E}(XY)]^2$ 得到 (3.5)，并且知道 (3.5) 中等号成立当且仅当 Σ 退化, 当且仅当有不全为零的常数 a, b 使 $\mathrm{E}(aX + bY)^2 = 0$, 当且仅当 (用例 2.8) 有不全为零的常数 a, b 使 $P(aX + bY = 0) = 1$.

4.4 协方差和相关系数

协方差和相关系数的定义

- 设 $\sigma_X = \sqrt{\sigma_{XX}}$, $\sigma_Y = \sqrt{\sigma_{YY}}$ 分别是 X, Y 的标准差.
- **定义 4.1** 设 $\mu_X = \mathrm{E}X$, $\mu_Y = \mathrm{E}Y$ 存在,

  (1) 当 $\mathrm{E}|(X-\mu_X)(Y-\mu_Y)| < \infty$, 称

  $$\mathrm{E}[(X-\mu_X)(Y-\mu_Y)] \tag{4.1}$$

  为随机变量 X, Y 的**协方差** (covariance), 记做 $\mathrm{Cov}(X,Y)$ 或 σ_{XY}. 当 $\mathrm{Cov}(X,Y)=0$ 时, 称 X, Y **不相关**.

  (2) 当 $0 < \sigma_X\sigma_Y < \infty$, 称

  $$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \tag{4.2}$$

  为 X, Y 的 **相关系数** (correlation coefficient). 有时也用 $\rho(X,Y)$ 表示相关系数 ρ_{XY}.
- 在定义 4.1 中, 引入 X, Y 的标准化

  $$\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y},$$

  则有

  $$\rho_{XY} = \mathrm{E}\Big[\big(\frac{X-\mu_X}{\sigma_X}\big)\big(\frac{Y-\mu_Y}{\sigma_Y}\big)\Big].$$
- 下面是计算协方差的常用公式.

$$\sigma_{XY} = \mathrm{E}(XY) - (\mathrm{E}X)(\mathrm{E}Y). \tag{4.3}$$

相关系数的性质

- 从定义 4.1 和内积不等式 (见定理 3.3) 马上得到相关系数的性质如下.
- **定理 4.1** 设 ρ_{XY} 是 X, Y 的相关系数, 则有

  1. $|\rho_{XY}| \le 1$,
  2. $|\rho_{XY}| = 1$ 的充分必要条件是有常数 a, b 使得

  $$P(Y = a + bX) = 1,$$

  3. 如果 X, Y 独立, 则 X, Y 不相关.
- 证明课外。

例 4.1

- (接例 2.4)
- 设 (X,Y) 在单位圆 $D=\{(x,y)|\,x^2+y^2\le 1\}$ 内均匀分布, 则 X, Y 不相关, 也不独立.
- **证** 由例 2.4 知道 $\mathrm{E}(X)=\mathrm{E}(Y)=0$, 于是

$$\mathrm{Cov}(X,Y)=\int\int_{R^2} xyf(x,y)dxdy=\frac{1}{\pi}\int_{-1}^{1} ydy\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x\,dx=0.$$

- 所以 X,Y 不相关.
- 又由 §3.3 的例 3.1 知道联合密度不能写成两个边缘密度的乘积, 所以 X,Y 不独立.

例 4.2

- 相关系数 ρ_{XY} 只表示了 X,Y 间的线性关系.
- 当 $\rho_{XY}=0$, 尽管称 X,Y 不相关, 它们之间还可以有很强的非线性关系.
- 例如当 $Y=X^2$, $X\sim N(0,\sigma^2)$ 时, (X,Y) 总在抛物线 $y=x^2$ 上, 但是 X,Y 不相关, 因为

$$\mathrm{Cov}(X,Y)=\mathrm{E}[X(Y-\sigma^2)]=\mathrm{E}X^3-\sigma^2\mathrm{E}X=0.$$

例 4.3

- (接 §3.3 例 3.6)
- 设 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$, 则 $\rho_{XY}=\rho$, 并且 X, Y 独立的充分必要条件是 X, Y 不相关.
- **证明：** 课外。

随机向量和随机矩阵的期望

- 设 $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$ 是随机向量，若 $\mathrm{E}X_i=\mu_i$，则记

$$\boldsymbol{\mu}=(\mu_1,\mu_2,\ldots,\mu_n)$$

称

$$\mathrm{E}\boldsymbol{X}=\boldsymbol{\mu}$$

- 若 $Y_{ij}(i=1,\ldots,m,\ j=1,\ldots,n)$ 是随机变量，$\mathrm{E}Y_{ij}$ 存在，记

$$\boldsymbol{Y}=\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & & \vdots \\ Y_{m1} & Y_{m2} & \cdots & Y_{mn} \end{pmatrix}$$

则称

$$\mathrm{E}\boldsymbol{Y}=\begin{pmatrix} \mathrm{E}Y_{11} & \mathrm{E}Y_{12} & \cdots & \mathrm{E}Y_{1n} \\ \mathrm{E}Y_{21} & \mathrm{E}Y_{22} & \cdots & \mathrm{E}Y_{2n} \\ \vdots & \vdots & & \vdots \\ \mathrm{E}Y_{m1} & \mathrm{E}Y_{m2} & \cdots & \mathrm{E}Y_{mn} \end{pmatrix}$$

随机向量和随机矩阵的性质

- 设 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 如上定义，$\mathrm{E}\boldsymbol{X}$ 和 $\mathrm{E}\boldsymbol{Y}$ 存在。
- 对任意常数向量 $\boldsymbol{a}=(a_1,\ldots,a_n)$, $k\times m$ 常数矩阵 A 和 $n\times l$ 常数矩阵 B，有

$$\begin{aligned} \mathrm{E}(a\boldsymbol{X}^T) =& a\mathrm{E}\boldsymbol{X}^T, \\ (\mathrm{E}\boldsymbol{Y})^T =& \mathrm{E}(\boldsymbol{Y}^T), \\ \mathrm{E}(A\boldsymbol{Y}) =& A\ \mathrm{E}(\boldsymbol{Y}), \\ \mathrm{E}(\boldsymbol{Y}B) =& \mathrm{E}(\boldsymbol{Y})\ B, \\ \mathrm{E}(A\boldsymbol{Y}B) =& A\ \mathrm{E}(\boldsymbol{Y})\ B. \end{aligned} \tag{4.6}$$

- 证明略。

协方差阵

- **定义 4.2** 如果随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的数学期望 $\boldsymbol{\mu}=\mathrm{E}\boldsymbol{X}$ 存在, 每个 X_i 的方差 $\mathrm{Var}(X_i)<\infty$, 就称

$$\Sigma=\mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{\mu})]=(\sigma_{ij})_{n\times n} \tag{4.7}$$

为 $\boldsymbol{X}$ 的**协方差矩阵**. 其中

$$\sigma_{ij}=\mathrm{Cov}(X_i,X_j) \tag{4.8}$$

是 X_i, X_j 的协方差.

- 协方差矩阵 Σ 是对称矩阵.
- 如果矩阵 A 的行列式 $\det(A)=0$, 就称 A 是退化的.

- **定理 4.2** 设 Σ 是 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的协方差矩阵, 则

  (1) Σ 是非负定矩阵,

  (2) Σ 退化的充分必要条件是有不全为零的常数 $a_1,a_2,\cdots,a_n$ 使得

$$P\Big(\sum_{i=1}^{n}a_i(X_i-\mu_i)=0\Big)=1. \tag{4.9}$$

  其中 $\mu_i=\mathrm{E}X_i$.

定理 4.2 证明

- 任取一个 n 维实向量 $\boldsymbol{a}=(a_1,a_2,\cdots,a_n)$, 有

$$\begin{aligned}\boldsymbol{a}\Sigma\boldsymbol{a}^{\mathrm{T}}=&\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j\sigma_{ij}\\ =&\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)]\\ =&\mathrm{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j(X_i-\mu_i)(X_j-\mu_j)\Big]\\ =&\mathrm{E}\big[\sum_{i=1}^{n}a_i(X_i-\mu_i)\big]^2\\ =&\mathrm{Var}\big[\sum_{i=1}^{n}a_i(X_i-\mu_i)\big]\\ \geq&0.\end{aligned} \tag{4.10}$$

- 所以 Σ 非负定.

- 从 (4.10) 看出, Σ 退化的充分必要条件是有非零向量 $\boldsymbol{a}$ 使得

$$\mathrm{Var}\big[\sum_{i=1}^{n}a_i(X_i-\mu_i)\big]=0.$$

  再用定理 3.1 的 (3) 得到本定理的 (2).

第五章　多元正态分布和极限定理

5.1　多元正态分布

多元正态分布

- 本节中的向量都是列向量, 用 A^{T} 表示矩阵 A 的转置.
- 设 $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)^{\mathrm{T}}$ 是 n 维常数列向量, B 是 $n \times m$ 常数矩阵, $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m$ 是相互独立都服从标准正态分布的随机变量, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m)^{\mathrm{T}}$.
- **定义 1.1**　如果

$$\boldsymbol{X} = \boldsymbol{\mu} + B\boldsymbol{\varepsilon}, \tag{1.1}$$

  就称随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 服从 n 元 (或多元) 正态分布, 记做 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, 其中 $\Sigma \triangleq BB^{\mathrm{T}}$.
- 上面定义的 $\boldsymbol{X}$ 的数学期望和方差阵为

$$\begin{aligned}
\mathrm{E}\boldsymbol{X} =& \mathrm{E}\left[\boldsymbol{\mu} + B\boldsymbol{\varepsilon}\right] = \boldsymbol{\mu} + B\mathrm{E}\boldsymbol{\varepsilon} \\
=& \boldsymbol{\mu} + B \cdot \mathbf{0} = \boldsymbol{\mu} \\
\mathrm{Var}(\boldsymbol{X}) =& \mathrm{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\right] \\
=& \mathrm{E}\left[(B\boldsymbol{\varepsilon})(B\boldsymbol{\varepsilon})^T\right] = \mathrm{E}\left[B(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)B^T\right] \\
=& B\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)B^T = BI_mB^T = BB^T = \Sigma
\end{aligned}$$

- **注意:**　$\boldsymbol{\mu}$ 和 Σ 完全决定多元正态分布的随机向量 $\boldsymbol{X}$ 的联合分布。

多元正态分布与一元正态分布

- 对 $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ 独立同标准正态分布，记 $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)^T$，$\boldsymbol{a} = (a_1, a_2, \ldots, a_m)^T$ 是 m 维常数向量，μ 为实数，按照定义 1.1，$X = \mu + \boldsymbol{a}^T\boldsymbol{\varepsilon}$ 是 $n = 1$ 的多元正态分布，X 是否服从一元正态分布？

- 显然 $EX=\mu$, $\mathrm{Var}(X)=\sum_{i=1}^m a_i^2=\boldsymbol{a}^T\boldsymbol{a}$。
- 如果 $\boldsymbol{a}$ 的元素不全为零，则 $\mathrm{Var}(X)>0$，这时 $X\sim \mathrm{N}(\mu,\boldsymbol{a}^T\boldsymbol{a})$。多元正态分布在 $n=1$ 且 $\Sigma>0$ 时与一元正态分布相同。

证明

- 用数学归纳法。当 $m=1$ 时，$X=\mu+a_1\varepsilon_1$，当 $a_1\neq 0$ 时已证明 $X\sim \mathrm{N}(\mu,a_1^2)$。
- 设结论对 m 成立，要证明 $X=\mu+a_1\varepsilon_1+\cdots+a_m\varepsilon_m+a_{m+1}\varepsilon_{m+1}$ 服从 $\mathrm{N}(\mu,\sum_{i=1}^{m+1}a_i^2)$。
- 如果 $(a_1,\ldots,a_m)=(0,\ldots,0)$，则 $X=\mu+a_{m+1}\varepsilon_{m+1}\sim \mathrm{N}(\mu,a_{m+1}^2)$。如果 $a_{m+1}=0$, 则由归纳法假设 $X=\mu+a_1\varepsilon_1+\cdots+a_m\varepsilon_m\sim \mathrm{N}(\mu,\sum_{i=1}^m a_i^2)$。
- 设 $(a_1,\ldots,a_m)$ 不全为零，$a_{m+1}\neq 0$。令 $Z=\mu+a_1\varepsilon_1+\cdots+a_m\varepsilon_m$, 则 Z 与 ε_{m+1} 相互独立，由归纳法假设可知 $Z\sim \mathrm{N}(\mu,\sum_{i=1}^m a_i^2)$。
- $X=Z+a_{m+1}\varepsilon_{m+1}$。
- 因为 Z,ε_{m+1} 独立，所以 (Z,ε_{m+1}) 有联合密度

$$f(z,u)=(2\pi)^{-1}\left(\sum_{i=1}^m a_i^2\right)^{-1/2}\exp\left\{-\frac{1}{2}\left[\frac{(z-\mu)^2}{\sum_{i=1}^m a_i^2}+u^2\right]\right\}$$

- 做变换

$$\begin{pmatrix} X\\ Y\end{pmatrix}=\begin{pmatrix}1 & a_{m+1}\\ 0 & 1\end{pmatrix}\begin{pmatrix} Z\\ \varepsilon_{m+1}\end{pmatrix}$$

- 此变换是一一变换，逆变换为

$$\begin{pmatrix} Z\\ \varepsilon_{m+1}\end{pmatrix}=\begin{pmatrix}1 & -a_{m+1}\\ 0 & 1\end{pmatrix}\begin{pmatrix} X\\ Y\end{pmatrix}$$

逆变换的 Jacobi 行列式等于 1。

所以，(X,Y) 的联合密度为

$$\begin{aligned}f(x,y)=&(2\pi)^{-1}\left(\sum_{i=1}^m a_i^2\right)^{-1/2}\\ &\exp\left\{-\frac{1}{2}\left[\frac{(x-\mu-a_{m+1}y)^2}{\sum_{i=1}^m a_i^2}+y^2\right]\right\}\\ =&c_1\exp\left\{-\frac{1}{2}\left[\frac{(x-\mu-a_{m+1}y)^2}{\sum_{i=1}^m a_i^2}+y^2\right]\right\}\end{aligned}$$

X 的边缘密度为

$$
\begin{aligned}
&f_X(x)\\
=&c_1\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^{m+1}a_i^2}{\sum_{i=1}^{m}a_i^2}y^2-2\frac{a_{m+1}}{\sum_{i=1}^{m}a_i^2}(x-\mu)y+\frac{(x-\mu)^2}{\sum_{i=1}^{m}a_i^2}\right]\right\}dy\\
=&c_1\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sum_{i=1}^{m}a_i^2}\right\}\\
&\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^{m+1}a_i^2}{\sum_{i=1}^{m}a_i^2}\left[y^2-2\frac{a_{m+1}}{\sum_{i=1}^{m+1}a_i^2}(x-\mu)y\right]\right\}dy\\
=&c_1\exp\left\{-\frac{1}{2}\left[\frac{1}{\sum_{i=1}^{m}a_i^2}-\frac{a_{m+1}^2}{(\sum_{i=1}^{m}a_i^2)\left(\sum_{i=1}^{m+1}a_i^2\right)}\right](x-\mu)^2\right\}\\
&\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^{m+1}a_i^2}{\sum_{i=1}^{m}a_i^2}\left[y-\frac{a_{m+1}}{\sum_{i=1}^{m+1}a_i^2}(x-\mu)\right]^2\right\}dy\\
=&c_2\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sum_{i=1}^{m+1}a_i^2}\right\}
\end{aligned}
$$

其中 c_2 是使得 $\int_{-\infty}^{\infty}f_X(x)\,dx=1$ 的常数，因为 $f_X(x)$ 是密度，所以 $c_2=(2\pi)^{-1/2}\left(\sum_{i=1}^{m+1}a_i^2\right)^{-1/2}$。于是 $X=\mu+\sum_{i=1}^{m+1}a_i\varepsilon_i\sim \mathrm{N}(\mu,\sum_{i=1}^{m+1}a_i^2)$。证毕。

多元正态分布的性质

- **性质 1** 多元正态分布的边缘分布也是多元 (或一元) 正态分布.
- **证:** 这是因为，$\boldsymbol{X}=\boldsymbol{\mu}+B\boldsymbol{\varepsilon}$ 则其一部分分量可以写成用 $\boldsymbol{\mu}$ 和 B 的对应行线性组合的结果。
- **性质 2** 如果 $\boldsymbol{X}\sim N(\boldsymbol{\mu},\Sigma)$, A 是 $m\times n$ 矩阵, 则

$$\boldsymbol{Y}=A\boldsymbol{X}\sim N(A\boldsymbol{\mu},A\Sigma A^{\mathrm{T}}).$$

- 特别对向量 $\boldsymbol{a}=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$, 有

$$\sum_{j=1}^{n}a_jX_j=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\sim \mathrm{N}(\boldsymbol{a}^{\mathrm{T}}\mu,\boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a}).$$

- **证** 由 $\boldsymbol{Y}=A\boldsymbol{X}=(A\boldsymbol{\mu})+(AB)\boldsymbol{\varepsilon}$ 和定义 1.1 得到结论.

- **性质 3** 设 $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$. 当 $\Sigma = BB^{\mathrm{T}}$ 正定时, $\boldsymbol{X}$ 是连续型随机向量, 有联合密度:

$$f(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(\Sigma)}} \exp\Big[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big],$$
$$\boldsymbol{x} = (x_1, x_2, \cdots x_n)^{\mathrm{T}} \in R^n. \tag{1.3}$$

- 证明略。

- **性质 4** 如果 Σ 正定，$\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, 则 $X_1, X_2, \ldots, X_n$ 相互独立的充分必要条件是

$$\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$$

- 证明：利用性质 3 以及独立的充分必要条件是联合密度等于边缘密度乘积。

- **性质 5** 设 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, 则 $X_1, X_2, \cdots, X_n$ 相互独立都服从标准正态分布的充分必要条件是 $\boldsymbol{\mu} = 0, \Sigma = I$, 即 $\boldsymbol{X} \sim N(0, I)$.

- **证明:** 必要性: 设各分量 iid 标准正态分布，则

$$\boldsymbol{X} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + I_n \boldsymbol{\varepsilon}$$

  其中 $\varepsilon = (X_1, \ldots, X_n)^T$, 由定义可知 $\boldsymbol{X} \sim \mathrm{N}(\mathbf{0}, I_n)$.

- 充分性: 若 $\boldsymbol{\mu} = \mathbf{0}$, $\Sigma = I_n$，由性质 1 可知各 X_i 为正态分布，且为 N(0,1) 的正态分布，再由性质 4 可知 $X_1, X_2, \ldots, X_n$ 相互独立。

- **性质 6** 设 $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, 有可逆矩阵 A 使得 $\Sigma = AA^{\mathrm{T}}$, 则

$$\boldsymbol{Y} = A^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim N(\mathbf{0}, I).$$

- **证** 由性质 2 知 $\boldsymbol{Y}$ 服从多元正态分布, 数学期望和方差阵为

$$\begin{aligned} \mathrm{E}\boldsymbol{Y} =& A^{-1}\mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu}) = 0, \\ \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}) =& A^{-1}\mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}]A^{-\mathrm{T}} \\ =& A^{-1}AA^{\mathrm{T}}A^{-\mathrm{T}} = I. \end{aligned}$$

例 1.1

- 设 $X_1, X_2, \cdots, X_n$ 相互独立, $X_j \sim N(\mu_j, \sigma_j^2)$, 则对非零向量 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}}$, $Y = \sum_{j=1}^n a_j X_j \sim \mathrm{N}\big(\sum_{j=1}^n a_j\mu_j,\ \sum_{j=1}^n a_j^2\sigma_j^2\big)$.
- **证** 定义 $\varepsilon_j = (X_j - \mu_j)/\sigma_j$, 则 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 相互独立都服从标准正态分布, 并且 $X_j = \mu_j + \sigma_j\varepsilon_j$.
- 从定义 1.1 知道

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

  服从多元正态分布.
- 从性质 2 知道 $Y = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}$ 服从正态分布.
- 容易计算 $\mathrm{E}Y = \sum_{j=1}^n a_j\mu_j$, $\mathrm{Var}(Y) = \sum_{j=1}^n a_j^2\sigma_j^2$, 故结论成立.

例 1.2

- 如果 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, 其中 $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^{\mathrm{T}}$, $\Sigma = (\sigma_{ij})_{n\times n}$. 则 $X_j \sim N(\mu_j, \sigma_{jj})$.
- **证** 设向量 $\boldsymbol{a} = (0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}$ 的第 j 个元素是 1, 其余是零. 则 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu} = \mu_j$, $\boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a} = \sigma_{jj}$. 从性质 2 得到 $X_j = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \sim \mathrm{N}(\mu_j, \sigma_{jj})$.

5.2 大数律

概率频率定义与大数律

- 在 n 次独立重复试验中, 引入

$$X_j = \begin{cases} 1, & \text{当第 } j \text{ 试验成功,} \\ 0, & \text{当第 } j \text{ 试验不成功.} \end{cases}$$

  则 $S_n = X_1 + X_2 + \cdots + X_n$ 是 n 次试验中的成功次数, 由概率的频率定义知道, 对于成功的频率 $\overline{X}_n = S_n/n$, 有

$$\lim_{n\to\infty} \overline{X}_n = P(X_1 = 1) = \mathrm{E}X_1. \tag{2.1}$$

- 下面的强大数定律将 (2.1) 进行了推广.

依概率收敛

- 称随机变量的序列 $\{\xi_n\} = \{\xi_1, \xi_2, \cdots\}$ 为**随机序列** (random sequence).
- **定义 2.1** 设 $\{\xi_n\}$ 是随机序列, ξ 是随机变量. 如果对任何 $\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \epsilon) = 0, \tag{2.2}$$

  就称 ξ_n 依概率收敛到 ξ, 记做 $\xi_n \xrightarrow{\mathrm{p}} \xi$.
- 其含义是 n 很大时, ξ_n 与 ξ 有非零差距的可能性很小。

弱大数律

- **定理 2.1** 设随机序列 $\{X_n\}$ 独立同分布, 如果 $\mu = \mathrm{E}X_1$ 有限, 则有

$$\overline{X}_n \xrightarrow{\mathrm{p}} \mu \quad (n \to \infty) \tag{2.5}$$

  其中

$$\overline{X}_n \triangleq \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j.$$

- **证** 只对 $\sigma^2 = \mathrm{Var}(X) < \infty$ 的情况证明. 因为 $\overline{X}_n$ 分别有数学期望和方差

$$\mathrm{E}\bar{X}_n = \mu, \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n},$$

  利用马尔柯夫不等式 (见 §4.3 定理 3.2) 得到, 对任何 $\epsilon > 0$,

$$\begin{aligned} &P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)\Big| \geq \epsilon\Big) \\ \leq &\frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0,\ n \to \infty. \end{aligned}$$

- 通常把类似于 (2.5) 的结论称为弱大数律 (weak law of large numbers).

例 2.1

- (接 §4.1 的例 1.4) 在赌对子时, 甲每次下注 100 元. 如果他连续下注 n 次, 证明他的盈利 S_n 满足

$$P(S_n \leq -18n) \to 1.$$

- **证** 用 X_i 表示甲第 i 次下注的盈利, 则 $X_1, X_2, \cdots, X_n$ 独立同分布. 由 §4.1 的例 1.4 知道 $\mu = \mathrm{E}X_i = -18.6$. $S_n = X_1 + X_2 + \cdots + X_n$.

- 利用

$$\begin{aligned}\{S_n > -18n\} =& \{\bar{X}_n - \mu > -18 + 18.6\} \\ =& \{\bar{X}_n - \mu > 0.6\} \\ \subset& \{|\bar{X}_n - \mu| > 0.6\},\end{aligned}$$

  和定理 2.1 得到, $n \to \infty$ 时,

$$P(S_n > -18n) \leq P(|\bar{X}_n - \mu| > 0.6) \to 0.$$

- 于是

$$P(S_n \leq -18n) = 1 - P(S_n > -18n) \to 1.$$

- 说明下注的次数 n 越多, 至少输 $18n$ 元的概率越大.

强大数律

- **定义 2.2** 设 $\{\xi_n\}$ 是随机序列, ξ 是随机变量. 如果

$$P\Big(\lim_{n\to\infty} \xi_n = \xi\Big) = 1,$$

  就称 ξ_n 以概率 1 收敛到 ξ, 记做 $\xi_n \to \xi$, wp1，或 a.s.

- **定理 2.2** 如果 $\{X_j\}$ 是独立同分布的随机序列, $\mu = \mathrm{E}X_1$, 则

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu, \text{ wp1.} \tag{2.6}$$

- (2.6) 的另一个表达方式是

$$P\big(\lim_{n\to\infty} \bar{X}_n = \mu\big) = 1.$$

- 类似于 (2.6) 的结果称为**强大数律** (strong law of large numbers). 从强大数律结论 (2.6) 知道概率的频率定义是合理的.

- 强大数律结论比弱大数律结论要强:

- **定理 2.3** 如果 $\xi_n \to \xi$, wp1., 则 $\xi_n \xrightarrow{\mathrm{p}} \xi$.

例 2.2

- 在多次独立重复试验过程中, 小概率事件必然发生.
- **证:** 设 p 是任意小的正数, 事件 $A_1, A_2, \ldots$ 相互独立, $P(A_i) = p$. 用 $\mathrm{I}[A_i]$ 表示 A_i 的示性函数, 则 $\mathrm{I}[A_i]$ 独立同分布.
- 由强大数律得到
$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{I}[A_i] \to p, \text{ wp1}.$$
- 所以
$$\sum_{i=1}^{\infty}\mathrm{I}[A_i] = \infty, \text{ wp1}.$$
说明有无穷个 A_i 发生的概率是 1.

5.3 中心极限定理

中心极限定理—问题

- 强大数律和弱大数律分别讨论了随机序列部分和的依概率收敛和以概率 1 收敛.
- 中心极限定理讨论对充分大的 n, 随机变量的部分和
$$X_1 + X_2 + \cdots + X_n$$
的概率分布问题.

例 3.1: 二项分布

- 独立地重复某一试验，设
$$X_i = \begin{cases} 1 & \text{试验成功} \\ 0 & \text{试验失败} \end{cases} \qquad i = 1, 2, \ldots, n$$
则 X_i iid $\sim \mathrm{B}(1, p)$(两点分布)。
- 令
$$S_n = X_1 + X_2 + \cdots + X_n$$
则 S_n 为 n 次独立试验中成功的次数，$S_n \sim \mathrm{B}(n, p)$。
- 从演示看出 $n \to \infty$ 时 S_n 的分布形状很象正态分布。

例 3.2: Poisson(泊松) 分布

- 若 $\{X_j\}$ iid P(λ), 则由 §3.4 的例 4.1 知道部分和

$$S_n = \sum_{j=1}^{n} X_j \sim \mathrm{P}(n\lambda).$$

- 从演示看出 $n \to \infty$ 时 S_n 的分布形状很象正态分布。

例 3.4: 几何分布的部分和

- 设 $\{X_j\}$ 独立同分布都服从几何分布 $P(X = k) = pq^{k-1}$, $k = 1, 2, \cdots$, $p + q = 1$.
- 可以将 $S_n = \sum_{j=1}^{n} X_j$ 设想成第 n 次击中目标时的射击次数 (参考几何分布的背景), 于是得到

$$P(S_n = k) = C_{k-1}^{n-1} p^n q^{k-n}, \ k = n, n+1, \ldots.$$

上述分布称为**帕斯卡分布**.

- 从演示看出 S_n 的概率分布当 $n \to \infty$ 时越来越接近于正态分布。
- 注：得到第 n 次成功前失败的次数 Y 的分布称为负二项分布，易见

$$P(Y = k) = C_{n+k-1}^{k} p^n (1-p)^k, \qquad k = 0, 1, 2, \ldots$$

且 $S_n = Y + n$。

例 3.4: 指数分布的独立和

- 设 $\{X_j\}$ 独立同分布都服从指数分布 E(λ), 可以证明部分和 $S_n = \sum_{j=1}^{n} X_j$ 服从 $\Gamma(n, \lambda)$ 分布.
- 从演示看出 S_n 的密度趋向于正态分布。

中心极限定理

- **定理 3.1** (中心极限定理) 设随机序列 $\{X_j\}$ 独立同分布, 有共同的数学期望 μ 和方差 σ^2. 部分和由 $S_n = \sum_{j=1}^{n} X_j$ 定义, 则 S_n 的标准化

$$\xi_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

依分布收敛到标准正态分布. 即对任何 x,

$$\lim_{n\to\infty} P(\xi_n \le x) = \Phi(x). \tag{3.2}$$

这里 $\Phi(x)$ 是标准正态分布的分布函数.

- 我们把结论 (3.2) 记成 $\xi_n \xrightarrow{\mathrm{d}} N(0,1)$, 其中的 d 表示依分布收敛.
- 中心极限定理的应用: 可以用 N(0,1) 近似计算关于 ξ_n 的概率，用 $\mathrm{N}(n\mu, n\sigma^2)$ 近似计算关于 S_n 的概率。
- 中心极限定理对大数律的补充：$\bar{X}_n \to \mu$, 收敛速度如何？因为 $\xi_n = \sigma^{-1}\sqrt{n}(\bar{X}_n - \mu)$ 渐近分布为 N(0,1), 所以 $|\bar{X}_n - \mu|$ 趋于零的速度是 $\frac{1}{\sqrt{n}}$。

例 3.6: 近似计算

- 当辐射的强度超过每小时 0.5 毫伦琴 (mr) 时, 辐射会对人的健康造成伤害. 设一台彩电工作时的平均辐射强度是 0.036(mr/h), 方差是 0.0081. 则家庭中一台彩电的辐射一般不会对人造成健康伤害. 但是彩电销售店同时有多台彩电同时工作时, 辐射可能对人造成健康伤害. 现在有 16 台彩电同时工作, 问这 16 台彩电的辐射量可以对人造成健康伤害的概率.
- **解** 用 X_i 表示第 i 台彩电的辐射量 (mr/h), 则 X_i 的数学期望是 $\mu = 0.036$, 方差是 $\sigma^2 = 0.0081$. $S_n = X_1 + X_2 + \cdots + X_{16}$ 是 $n = 16$ 台彩电的辐射量. 题目要求 $P(S_n > 0.5)$.
- 认为 $\{X_i\}$ 独立同分布时, 按照定理 3.1,

$$\xi_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

近似服从 $N(0,1)$ 分布, 于是

$$\begin{aligned} P(S_n > 0.5) =& P\Big(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} > \frac{0.5 - n\mu}{\sqrt{n\sigma^2}}\Big) \\ =& P\Big(\xi_n > \frac{0.5 - 16 \times 0.036}{\sqrt{16 \times 0.0081}}\Big) \\ =& P(\xi_n > -0.211) \\ \approx& \Phi(0.211) = 0.58. \end{aligned}$$

这 16 台彩电以大约 58% 的概率会对人造成健康伤害.

二项分布的正态近似

- **推论 3.3** 设 $S_n \sim B(n,p)$, $p = 1 - q \in (0,1)$, 则

$$\frac{S_n - np}{\sqrt{npq}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1). \tag{3.3}$$

- **证明** 设 $\{X_n\}$ 是独立同分布的随机序列, X_i 服从两点分布 $B(1,p)$. 则 $\xi_n = X_1 + X_2 + \cdots + X_n$ 和 S_n 同分布, $\mathrm{E}\xi_n = np$, $\mathrm{Var}(\xi_n) = npq$.

- 于是

$$\frac{S_n - np}{\sqrt{npq}} \text{ 和 } \frac{\xi_n - np}{\sqrt{npq}}$$

同分布.

- 由定理 3.1 知道当 $n \to \infty$, 对 $x \in (-\infty, \infty)$,

$$P\Big(\frac{S_n - np}{\sqrt{npq}} \le x\Big) = P\Big(\frac{\xi_n - np}{\sqrt{npq}} \le x\Big) \to \Phi(x).$$

例 3.7: 用正态分布计算二项分布

- 设 $S_n \sim \mathrm{B}(n,p)$ 则 S_n 近似 $\mathrm{N}(np,\, npq)$ 分布。设 $X \sim \mathrm{N}(np, npq)$, 设 a, b 为非负整数。

- 由中心极限定理 n 较大时

$$p = \Pr(a \le S_n \le b) \approx P(a < X < b) \qquad (*)$$

- 但是注意 S_n 是取整数值的，所以

$$p = \Pr(a \le S_n \le b) = \Pr(a - 1 < S_n < b + 1)$$

上式右端用正态近似和 (*) 不同。

- 为此取折衷，令

$$\begin{aligned} p =& \Pr(a \le S_n \le b) \approx \Pr(a - 0.5 < X < b + 0.5) \\ =& \Phi\Big(\frac{b + 0.5 - np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{a - 0.5 - np}{\sqrt{npq}}\Big). \end{aligned} \qquad (3.4)$$

称为连续性校正。此近似公式应在 n 充分大时使用，实际规则可以用 $\min(np, nq) > 5$。

- 特别地，

$$\Pr(S_n = a) \approx \Pr(a - 0.5 < X < a + 0.5).$$

例 3.8

- 某药厂试制了一种新药, 声称对贫血的治疗有效率达到 80%. 医药监管部门准备对 100 个贫血患者进行此药的疗效试验, 若这 100 人中至少有 75 人用药有效, 就批准此药的生产. 如果该药的有效率确实达到 80%, 此药被批准生产的概率是多少?

- **解** 用 S_n 表示这 $n(=100)$ 个患者中用药后有效的人数. 如果该药的有效率确实是 $p=80\%$, 则 $S_n \sim B(n,p)$.

- 由 $100p=80>5, 100(1-p)=20>5$, 知道可用近似公式 (3.4). 于是

$$\begin{aligned} P(\text{药被批准}) =& P(S_n \geq 75) \\ =& P(S_n > 74.5) \\ =& P\Big(\frac{S_n - np}{\sqrt{np(1-p)}} > \frac{74.5-np}{\sqrt{np(1-p)}}\Big) \\ =& P\Big(\frac{S_n - np}{\sqrt{np(1-p)}} > \frac{74.5-80}{\sqrt{80\times 0.2}}\Big) \\ \approx& 1-\Phi(-5.5/4) = \Phi(1.375) = 0.92. \end{aligned}$$

  于是药获得批准的概率是 92%. 如果有效率 $p>80\%$, 则获得批准的概率 $>92\%$(参考习题 7.29).

第六章　描述性统计

6.1　总体和参数

什么是统计学

- 统计学研究如何收集数据、分析数据、从数据做出有依据的推断结果。一言以蔽之，统计学是研究数据的科学。
- 统计学主要的数学工具是概率论，也广泛使用现代信息技术作为支撑，通过计算机和信息网络获取数据、进行建模、数据分析计算。
- 统计学是一门科学，不再是数学的一个分支。

描述性统计

- 统计学的做法分为两种：
  - 描述性统计:
  - 从数据样本中计算一些平均值、标准差、最小值、最大值等概括统计量，画直方图、散点图等描述图形。
  - 推断性统计：
  - 假定要研究的对象服从某种概率模型，收集数据后把数据用模型解释，并做出有概率意义的结论。

总体、个体和均值

- 所要调查的对象全体叫做**总体** (population), 总体中每个成员叫做**个体**。
- 举例：一批橘子甜不甜？这批橘子作为总体；或者，其甜度作为总体。
- 总体一般用随机变量作为数学模型。如：一批橘子的甜度分布情况用随机变量作为数学模型。

- **总体参数**是描述总体特性的指标，简称**参数**。
- 如果总体中的个体是有限个，称个体总数 N 为**总体容量**。
- **总体平均**或**总体均值**是参数。常用 μ 表示。如果知道总体的全部个体 (比如，某小学所有一年级新生的身高) $y_1, y_2, \ldots, y_N$ 则

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$$

- **总体方差**是参数。常记为 σ^2。如果知道总体的全部个体 $y_1, y_2, \ldots, y_N$ 则

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu)^2$$

- σ 称为**总体标准差**。

样本与估计

- 如果总体只有有限个样本虽然可以测量所有样本计算总体参数，但可能会消耗过大。
- 有些总体有无限个个体，比如，对某放射性物质测量固定长度时间内放射出的粒子数，每试验一次就有一个不同结果。
- 为了得到总体的信息，可以从总体中抽取一个**有代表性**的个体的集合，称为总体的一个**样本**。也叫观测数据。样本中个体的个数叫做**样本量** (sample size)。
- 试图用样本的情况去判断总体的情况。注意，“有代表性”是一个不容忽视的要求。
- 从总体中抽取样本的工作叫做抽样 (sampling)。
- 设一个样本为 $x_1, x_2, \ldots, x_n$, 可计算
- 样本均值

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- 样本方差

$$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

- $s = \sqrt{s^2}$ 称为样本标准差。

估计

- 如果样本确实是有代表性的，则当样本量 n 较大时，从样本计算的样本均值和样本方差可以与相应的总体均值和总体方差很接近。
- 利用样本计算出的对总体参数的估计值称为**估计** (estimator 或 estimate)。
- 不同的方法可能给出不同的估计，而评判估计优劣的标准也不是唯一的。这方面有一些数学理论。

例 1.1

- 实际问题中, 总体的样本量往往是非常大的, 这时从数据本身无法看清总体的情况. 样本均值和样本方差可以提供必要的信息.
- 比赛中甲, 乙两位射击运动员分别进行了 10 次射击, 成绩分别如下:

甲	9.5	9.9	9.9	9.9	9.8	9.7	9.5	9.3	9.6	9.6
乙	9.4	9.3	9.5	9.0	9.1	9.8	9.7	9.5	9.3	9.4

  问哪个运动员平均水平高, 哪个运动员水平更稳定.
- **解** 用 $\bar{x}$, s_x 和 $\bar{y}$, s_y 分别表示甲和乙成绩的样本均值和样本标准差.
- $$\bar{x} = 9.67,\ s_x = 0.2058, \quad \bar{y} = 9.4,\ s_y = 0.2449.$$
- 甲的平均水平和稳定性都比乙好.
- 知道样本方差后, 可以作出更好的比较结果.

6.2 抽样调查方法

抽样调查的可行性

- 抽样的可行性：汤的例子
  - 样本的随机性 (代表性)。如：把汤搅拌均匀。
  - 适当的样本量。不用太多也不能太少。
  - 样本量不必随总体增大而增大。
- 在大数据背景下，统计数据又有了盲人摸象的窘境: 数据多而繁杂，特征过多造成数据分布不集中。

抽样调查的必要性

- 为了从样本推断总体的情况，样本的代表性是最关键的问题。
- 调查全部总体不现实或不必要，如: 寿命试验。
- 抽样调查因为工作量较小, 控制可以更严格，所以有时比普查可以更准确。
- 但是，如何确保抽取的样本有代表性？

随机抽样

- 如果总体中的每个个体都有相同的机会被抽中，就称这样的抽样方法为**随机抽样**方法。
- 简单地分，抽样分为有放回抽取和无放回抽取。
- **无放回随机抽样**指在总体中随机抽出一个个体后, 下次在余下的个体中再进行随机抽样.
- **有放回随机抽样**指抽出一个个体, 记录下抽到的结果后放回, 摇匀后再进行下一次随机抽样.
- 无放回抽取从实现上和从精度上更好，总体容量 N 很大时两者差异很小。
- 提高样本量可以提高估计精度，但不是总体越大，考虑的特征越多，样本量也需要随之增大。

例 2.1

- 设 N 件产品中有 M 件次品, N, M 都是未知的. 用随机抽样方法估计这批产品的次品率 $p = M/N$.
- **解** 无放回地从中依次取 n 件, 用 Y 表示取得的次品数, 则 $Y \sim \mathrm{H}(n, M, N)$。
- $$\mathrm{E}Y = np, \ \ \mathrm{Var}(Y) = np(1-p)\frac{N-n}{N-1}.$$
- 用样本次品率 $\hat{p} = Y/n$ 估计 p 时, 有

$$\mathrm{E}\hat{p} = p, \ \ \mathrm{Var}(\hat{p}) = \frac{1}{n}p(1-p)\frac{N-n}{N-1}. \tag{2.1}$$

- 如果采用有放回的随机抽样, 用 X 表示取得的次品数, 则 $X \sim B(n,p)$。
- $$\mathrm{E}X = np, \ \ \mathrm{Var}(X) = np(1-p).$$
- 用这时的样本次品率 $\tilde{p} = X/n$ 估计 p 时, 有
$$\mathrm{E}\tilde{p} = p, \ \ \mathrm{Var}(\tilde{p}) = \frac{1}{n}p(1-p). \tag{2.2}$$
- $\mathrm{E}\hat{p} = \mathrm{E}\tilde{p} = p$, 说明这两种方法都是较好的估计方法, 没有偏差.
- 由于样本方差 $\mathrm{E}(Y/n-p)^2$ 描述的是 Y/n 向真实参数 p 的集中程度, 因而是描述估计精度的量.
- 样本方差越小, 说明估计的精度越高.
- $\mathrm{Var}(\hat{p}) < \mathrm{Var}(\tilde{p})$ 说明无放回随机抽样的估计精度更好。
- 但是当 N 比 n 大很多时, $(N-n)/(N-1)$ 接近于 1, 说明两种抽样方法差别不大.
- 另外 $\mathrm{Var}(\tilde{p})$ 与 N 无关, 说明要达到一定的估计精度, 只需要适当地增加 n.
- 并不是说总体数目 N 越大, 就需要多抽样.
- 无放回随机抽样下的情况也是类似的.
- 试验和理论都证明: 在随机抽样下, 样本均值 $\bar{x}$ 是总体均值 μ 很好的估计, 样本标准差 s 是总体标准差 σ 很好的估计.
- 在样本量不大时, 增加样本量可以比较好地提高估计的精确度.
- 例: 考虑某大学一年级 2000 个同学的平均身高 μ 时, 需要调查 50 同学的身高.
- 实现无放回的随机抽样的方法是先将 2000 个同学的学号分别写在 2000 张小纸片上, 然后放入一个大纸箱进行充分的摇匀. 最后从纸箱中无放回地抽取 50 个纸片. 纸片上的学号就是被选中的同学的学号.
- 现在一般采用计算机随机抽签方法。
- 以上问题的一种计算机抽签方法为: 对 2000 个学号 i,$i = 1, 2, \ldots, 2000$, 独立地从 U(0,1) 抽取 $U_i, i = 1, \ldots, 2000$ 对应到每个学生。从 $U_1, U_2, \ldots, U_{2000}$ 中选出最小的 50 个，其对应的学生作为样本。

随机抽样的无偏性

- 从总体 X 中等可能地随机抽取，不论是有放回还是无放回，得到的 $X_1, X_2, \ldots, X_n$ 看成随机变量，都可以证明

$$\mathrm{E}\bar{X} = \mu$$

- 样本在需要讨论其分布性质时看成随机变量,记做大写的 $X_1, X_2, \ldots, X_n$; 在讨论样本的具体取值时看成普通数值，记做小写的 $x_1, x_2, \ldots, x_n$。

代表性偏差失误举例

- **例 2.2** 1936 年是美国总统选举年. 这年罗斯福 (Roosevelt) 任美国总统期满, 参加第二届的连任竞选, 对手是堪萨斯州州长兰登 (Landon).
- 当时美国刚从经济大萧条中恢复过来, 失业人数仍高达 900 多万, 人们的经济收入下降了三分之一后开始逐步回升.
- 当时, 观察家们普遍认为罗斯福会当选.
- 而美国的“文学摘要”杂志的调查却预测兰登会以 57% 对 43% 的压倒优势获胜.
- “文学摘要”的预测是基于对二百四十万选民的民意调查得出的.
- 自 1916 年以来, 在历届美国总统的选举中“文学摘要”都做了正确的预测. “文学摘要”的威信有力地支持着它的这次预测.
- 但是选举的结果是罗斯福以 62% 对 38% 的压倒优势获胜.
- 此后不久“文学摘要”杂志就破产了.
- 要了解“文学摘要”预测失败的原因就必须检查他们的抽样调查方案.
- “文学摘要”是将问卷寄给了一千万个选民, 基于收回的 240 万份问卷得出的判断. 这些选民的地址是在诸如电话簿, 俱乐部会员名单等上查到的.
- 分析: 1936 年只有大约四分之一的家庭安装了电话. 由于有钱人才更有可能安装家庭电话和参加俱乐部, 所以“文学摘要”的调查方案漏掉了那些不属于俱乐部的穷人和没有安装电话的穷人, 这就导致了调查结果有排除穷人的偏向.

- 在 1936 年, 由于经济开始好转, 穷人普遍有赞同罗斯福当选的倾向, 富人有赞同兰登当选的倾向. “文学摘要”的调查结果更多地代表了富人的意愿, 导致了预测的失败.

- 抽样的方案应当公平的对待每一位选民和每一个群体, 以便得到选民的真实情况. 将哪一个群体排除在外的抽样方案都会导致有偏的样本, 从而导致错误的结论.

分层抽样方法

- 总体当中分为不同人群时 (如城镇和乡村)，虽然仍然进行等可能随机抽样，这样不同人群差异过大引起估计误差变大，而且操作也不方便。

- 好的作法是按人口比例在不同人群中分别进行随机抽样。

- 计算平均值等统计量时要用**加权**求和（平均）计算。

例 2.3

- 2000 年, 某市进行家庭年收入调查时, 分别对城镇家庭和农村家庭进行调查.

- 在全部城镇的 85,679 户中无放回随机抽取了 350 户, 在全部农村的 275,692 户中无放回随机抽取了 360 户.

- 调查结果如下:

- 城镇家庭年平均收入是 35612 元;

- 农村家庭年平均收入是 5623 元.

- 这里遇到了两个分总体 A_1 和 A_2, 第一个分总体 A_1 是所有城镇家庭的年收入, 第二个分总体 A_2 是所有农村家庭的年收入.

- 用 A 表示该市所有家庭的年收入时, 总体 A 是两个分总体 A_1 和 A_2 的并.

- 用 $\bar{x}_1$ 表示来自总体 A_1 的样本均值, 用 $\bar{x}_2$ 表示来自总体 A_2 的样本均值, 则

$$\bar{x}_1 = 35612,\ \bar{x}_2 = 5623.$$

- A_1 和 A_2 在 A 中所占的比例分别是

$$w_1 = \frac{85679}{85679+275692} = 0.2371. \qquad w_2 = \frac{275692}{85679+275692} = 0.7629.$$

- A 的总体均值 μ 的估计是

$$\bar{X} = w_1\bar{x}_1 + w_2\bar{x}_2 = 0.2371 \times 35612 + 0.7629 \times 5623 = 12733(\text{元}).$$

于是该市平均年家庭收入的估计是 12733 元.

- 上面的抽样调查问题中, 还可以把全部家庭再细分成城镇中的工人, 公务员, 教师等; 将农村家庭分成农民家庭, 农村干部家庭等.

一般分层抽样方法

- 把总体 A 分成 L 个互不相交子总体:

$$A = A_1 + A_2 + \cdots + A_L.$$

称这些子总体为层 (strata), 称 A_i 为第 i 层.

- 然后在每层中独立地进行随机抽样.

- 用 N 表示总体 A 的个体总数, 用 N_i 表示第 i 层的个体总数时, 有

$$N = N_1 + N_2 + \cdots + N_L.$$

- 我们称

$$w_i = \frac{N_i}{N},\ (i = 1, 2, \cdots, L)$$

为第 i 层的层权 (weight).

- 用 μ 表示 A 的总体均值.

- 对 $i = 1, 2, \cdots, L$, 用 n_i 表示从第 i 层抽出样本的个数，$\bar{x}_i$ 表示从第 i 层抽出样本的样本均值. 称

$$\bar{x}_{st} = w_1\bar{x}_1 + w_2\bar{x}_2 + \cdots + w_L\bar{x}_L$$

是总体均值 μ 的**简单估计**.

- 称

$$V(\bar{x}_{st}) \equiv w_1^2\mathrm{Var}(\bar{x}_1) + w_2^2\mathrm{Var}(\bar{x}_2) + \cdots + w_L^2\mathrm{Var}(\bar{x}_L)$$

是简单估计 $\bar{x}_{st}$ 的抽样方差.

- 抽样方差 $V(\bar{x}_{st})$ 是评价简单估计 $\bar{x}_{st}$ 的估计精度的指标. $V(\bar{x}_{st})$ 越小, 说明 $\bar{x}_{st}$ 越好.

- 当各层内总体方差相近时，各层样本量 n_i 应该正比于各层总体容量 N_i。

为什么要用加权平均

- 在例 2.3（城镇与农村收入）中，如果从城镇与农村都抽取相同的样本个数，采用直接两个平均值再简单地平均，而忽略实际农村人口数为城镇人口三倍以上的事实，就过于乐观地估计了所有人口的收入。

- 设所有城镇个体收入为 X_i, $i = 1, \ldots, N_1$, 所有农村个体收入为 Y_j, $j = 1, \ldots, N_2$.

- 则城镇总体收入平均值为

$$\mu_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} X_i;$$

  农村总体收入平均值为

$$\mu_2 = \frac{1}{N_2} \sum_{j=1}^{N_2} Y_j.$$

- 总体平均值为

$$\begin{aligned} \mu =& \frac{1}{N_1 + N_2} \left(\sum_{i=1}^{N_1} X_i + \sum_{j=1}^{N_2} Y_j \right) \\ =& \frac{1}{N_1 + N_2} (N_1 \mu_1 + N_2 \mu_2) \\ =& \frac{N_1}{N_1 + N_2} \mu_1 + \frac{N_2}{N_1 + N_2} \mu_2. \end{aligned}$$

- 设城镇样本平均值为 $\bar{x}$, 则 $\bar{x}$ 为 μ_1 的良好估计；设农村样本平均值为 $\bar{y}$, 则 $\bar{y}$ 为 μ_2 的良好估计。

- 为了估计总体平均值 μ，需要用加权公式

$$\hat{\mu} = \frac{N_1}{N_1 + N_2} \bar{x} + \frac{N_2}{N_1 + N_2} \bar{y}.$$

分层抽样方法的优点

- 同时得到分层的统计量。
- 容易保证样本代表性从而提高精度。
- 实施容易。

系统抽样

- 随机抽样有时难于实施，当个体排列本身比较随机时，根据某种固定规律抽取，也能达到类似随机抽样效果，称为**系统抽样**。

例 2.4

- 在调查某居民住宅区的 999 个住户对住宅区的环境满意程度时, 要按照 1 : 14 的比例进行抽样调查.
- 为方便抽样, 将这 999 户按门牌号码的顺序依次编号.
- 下面的每个数对应一户的门牌号码.

1	2	3	4	5	6	7	.....	13	14
15	16	17	18	19	20	21	.....	27	28
29	30	31	32	33	34	35	.....	21	22
..	..	..	..	..	..	..	.....	..	..
981	982	983	984	985	986	987	.....	993	994
995	996	997	998	999					

- 先在 1-14 中随机抽取一个数字, 如果抽到 7, 就调查排在第 7 列的所有家庭, 请这些家庭对小区环境的满意程度打分, 分数分为 $1, 2, 3, 4, 5$ 级.
- 第 7 列有 71 户, 所以样本量 $n = 71$.
- 这 71 户的平均分是样本均值.
- 用样本均值作为全体住户对小区环境的平均分的估计.
- 用 x_i 表示这 71 户中第 i 户的打分, 样本均值是

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_{71}}{71}.$$

- 称上面的抽样方法为系统抽样法.

系统抽样法的特点

- 如果总体中的个体按一定的方式排列, 在规定的范围内随机抽取一个个体, 然后按照制定好的规则确定其他个体的抽样方法称为**系统抽样法**.
- 最简单的系统抽样法是取得一个个体后, 按相同的间隔抽取其他个体.
- 系统抽样方法的主要优点是实施简单, 只需要先随机抽取第一个个体, 以后按规定抽取就可以了.
- 系统抽样法比随机抽样法简单, 随机抽样方法每次都要随机抽取个体.
- 如果了解总体中个体排列的规律, 设计合适的系统抽样规则可以增加估计的精度.

6.3 用样本估计总体分布

表格与图形概括

- 实际数据量可能很大，比如几千、几万、几十万、几百万观测值都是可能的。
- 直接浏览数据可以获得一些直观印象，但是不能形成总体分布概念。
- 总体分布包括：变量是离散取值还是连续取值的，如果离散取值，所有可取值集合是什么，每种取值出现多少次，占百分之几。
- 如果变量是连续取值的，需要了解变量的取值范围，然后在取值范围内分段，对每段的取值个数进行计数并计算百分比，可以画出每段的比例的图形（称为直方图），可以计算简单的样本平均值、标准差等，可以画密度估计图、茎叶图等。

频率分布表

- 对离散型总体（如性别、职业等），只要列出样本样本每个值的次数和比例。
- 例如，某班学生名单中，男生 30 个，女生 12 个，可以列出如下概括统计表格:

	人数	百分比
男生	30	71%
女生	12	29%

- 对于连续型总体，可以适当分组后列出每组的观测个数和百分比。
- 做出的表格称为频率分布表。
- 当样本量是 n, 可以参照下面的经验公式将数据分成大约

$$K = 1 + 4\lg n$$

段. 但是这里的经验公式只对分段起参考作用.

例 3.1

- 下面是某城市公共图书馆在一年中通过随机抽样调查得到的 60 天的读者借书数, 数据已经从小到大排列, 制作频率分布表.
- 数据:

213	230	239	289	291	301	308	310	311	312
318	318	337	343	344	348	349	351	360	362
368	372	374	379	383	385	390	393	396	399
400	404	406	425	429	430	436	438	440	441
444	446	450	453	456	458	471	473	475	483
484	495	498	498	521	524	549	556	568	584

- **解** 数据中的最小值是 213, 最大值是 584. 这 60 个数据就散布在闭区间 $[213, 584]$ 中.
- 取一个略大的区间 $(200, 600]$, 它的端点都是整数.
- 用经验公式计算出

$$K = 1 + 4\lg n = 1 + 4\lg 60 = 8.1126.$$

我们将 $(200, 600]$ 八等分, 排在下表的第一列.

- 计算出数据落入各段的个数 n_i, 填入第二列.
- 计算出数据落入各段的频率 (百分比)

$$f_1 = \frac{3}{60} = 5\%,\ f_2 = \frac{2}{60} \approx 3.3\%,\ \cdots,\ f_8 = \frac{3}{60} = 5\%,$$

依次填入第三列.

- 最后将各列之和填入最后一行, 得到如下频率分布表.

借出书数 i	发生次数 n_i	f_i = 发生频率
(200, 250]	3	5.0%
(250, 300]	2	3.3%
(300, 350]	12	20.0%
(350, 400]	14	23.3%
(400, 450]	12	20.0%
(450, 500]	11	18.3%
(500, 550]	3	5.0%
(550, 600]	3	5.0%
总计	60	99.9%

- 由于计算频率时四舍五入引起计算误差, 频率之和可能是 1 的近似.
- 从上述频率分布表可以方便地分析出以下结果:
- 有 8.3% 的工作日借出的图书少于等于 300 册;
- 有 63.3% 的工作日借出图书的数量在 301 至 450 册之间;
- 有 48.3% 的工作日借出的图书在 400 册以上;
- 只有 10% 的工作日借出的图书多于 500 册.
- 当总体 X 是全年每个工作日的借书数量时, 上述结果可以作为对总体的推测.

制作频率分布表的一般步骤

- **第一步**: 将数据从小到大排列, 将排列后的数据进行分段, 相等的数据必须分在同一段内.
- 每段中的数据被称为一组数据, 所以我们又把分段称为**分组**.
- 分段的多少应当适中.
- 分段过多, 数据过于分散, 不利于看出数据的特征和规律;
- 分段过少也不利于看到数据的特征和规律.
- **第二步**: 决定各段的长短. 在许多情况下, 为了方便, 除去第一和最后这两段, 可以把其他各段的长度取作相同.

- 还应当把各段的端点确定在便于记忆的数值上. 为了达到以上目的, 第一段的左端点可以比数据的最小值小一些, 最后一段的右端点可以比数据的最大值大一些.

- **第三步**: 绘制频率分布表的第一列 (参考例 3.1).

- **第四步**: 计算每段内数据的个数 n_i, 填入表格的第二列.

- **第五步**: 计算数据落在每一段内的频率 f_i, 填入表格的第三列.

- **第六步**: 将第 2,3 列之和填入最后一行形成总计.

- **注**: 由于频率分布表的制作没有统一的数据分段方法, 所以对相同的数据, 可以作出不同的频率分布表. 但是好的频率分布表应当是简单明了的.

频率直方图

- 离散型总体的各不同类型个数可以用条形图表示。

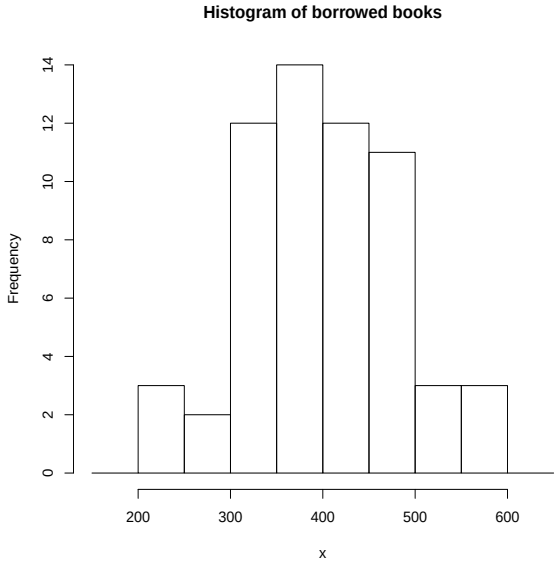
- 连续型数据分组后可以绘制频数直方图。这与频率分布表类似，只不过分组和频数都体现在图形中。以横坐标表示分组，以纵坐标表示频数，一个组用一个小矩形表示。

- 纵坐标也可以用频率，这样图形不变，只有纵坐标刻度变化。称为频率直方图。

- 纵坐标还可以适当伸缩使得小长方形的的总面积等于 1，用来作为分布密度估计，称为密度直方图。

- 下面的图用的是频数。

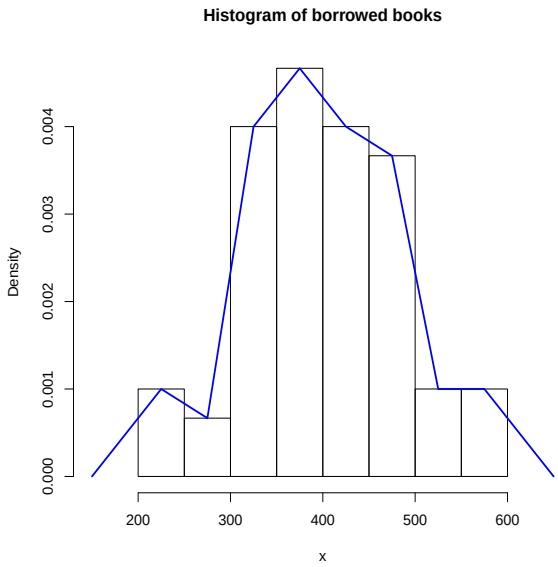
每日借出图书分布频数直方图



折线直方图

- 对密度直方图，在直方图最小值端和最大值端分别延伸出一个高度为零的小矩形，然后把两个相邻小矩形的顶部中点用折线连接起来，得到密度估计的折线图。
- 见下面图的蓝色线。

每日借出图书分布密度估计折线图



茎叶图

- 茎叶图可以看成水平放置的直方图。
- 茎叶图可以把所有数据点画到图上。
- 双茎叶图可以比较两个变量。

例 3.4

- 上海市 2004 年 7 月 10 日至 2004 年 7 月 31 日空气中可吸入颗粒物的监测数据：

85 85 66 71 62 52 55 59 52 62 59
70 80 96 97 94 62 51 57 67 96 93

- **解** 先将数据从小到大排列得到:

51 52 52 55 57 59 59
62 62 62 66 67
70 71
80 85 85
93 94 96 96 97

- 数据的十位上的数是 5,6,7,8,9, 把他们叫做“茎”, 排列下表的第一列;
- 茎 5 后面的个位数分别是 1,2,2,5,7,9,9, 把他们叫做茎 5 的“叶”, 排在茎 5 的后面, 按相同的方法把茎 6 的叶 2,2,2,6,7 排在茎 6 的右边, ……, 把茎 9 的叶 3,4,6,6,7 排在茎 9 的右边.
- 就得到了如下的茎叶图.

茎	叶
5	1225799
6	22267
7	01
8	055
9	34667

- 茎起到坐标轴的作用。
- 叶可以用来核查数据，叶的多少构成了横向的直方图。

- 从茎叶图中看出, 尽管这 22 天中可吸入颗粒物都是处于良的水平, 但是有较多的时间接近于优, 也有较多的时间接近于轻度污染.
- **注**: 可吸入颗粒物在 0 至 50 之间称为优; 可吸入颗粒物在 51 至 100 之间称为良; 可吸入颗粒物在 101 至 150 之间称为轻度污染.

每日借出图书样本的茎叶图

```
The decimal point is 2 digit(s) to the right of the |

2 | 134
2 | 99
3 | 0111122444
3 | 5556677788999
4 | 0000133344444
4 | 5556677888
5 | 00022
5 | 5678
```

数据值近似到了两位有效数字。

双茎叶图例子

- 上海市 2004 年 7 月 10 日至 2004 年 7 月 31 日空气中二氧化硫和二氧化氮的监测数据.
- 二氧化硫数据:

$$\begin{array}{lllllllllll} 55 & 62 & 54 & 71 & 60 & 51 & 55 & 56 & 51 & 58 & 61 \\ 62 & 69 & 73 & 72 & 69 & 58 & 42 & 42 & 65 & 77 & 73. \end{array}$$

- 二氧化氮数据:

$$\begin{array}{lllllllllll} 38 & 37 & 30 & 39 & 31 & 19 & 22 & 22 & 18 & 26 & 25 \\ 31 & 38 & 44 & 42 & 35 & 22 & 19 & 22 & 37 & 50 & 38. \end{array}$$

- 先将两组数据分别从小到大排列, 得到:
- 二氧化硫数据:

$$\begin{array}{llllllll} 42 & 42 & & & & & & \\ 51 & 51 & 54 & 55 & 55 & 56 & 58 & 58 \\ 60 & 61 & 62 & 62 & 65 & 69 & 69 & \\ 71 & 72 & 73 & 73 & 77 & & & \end{array}$$

- 二氧化氮数据:

  18 19 19
  22 22 22 22 25 26
  30 31 31 35 37 37 38 38 38 39
  42 44
  50.

- 两个茎叶图共享同一个茎作为坐标轴。
- 二氧化硫的叶子放在茎左侧，二氧化氮的叶子放在茎右侧。
- 图:

二氧化硫 树叶	树 茎	二氧化氮 树叶
	1	899
	2	222256
	3	0115778889
22	4	24
88655411	5	0
9952210	6	
73321	7	

- 纯以数值比较，二氧化硫的空气质量指标比二氧化氮的空气质量指标要差很多.
- 数据茎叶图的优点是利用了数据的每个信息, 从茎叶图中可以直观地看到数据的分布情况.
- 但是数据量很大时, 茎叶图的效果就不好了，因为这时的茎叶图会很长或很宽.

6.4 众数和中位数

众数和中位数

- 数据的频率分布表, 频率分布直方图和茎叶图都可以展示出数据的分布形状, 从中可以对数据有一个大致的了解.
- 需要更简单概括的描述方式。
- 平均值、标准差、众数、中位数等称为数据的数字特征。

众数

- 众数是观测值中出现次数最多的数，记为 M_0。
- 如果次数最多的不止一个，众数允许有多个。
- 众数有一定代表性。它受数据中极大或极小值变化的影响较小. 从分布的角度看, 众数出现的频率最高.
- 但是，众数的位置偶然性也比较大，现代采用较少。

例 4.1

- 某超市用随机抽样的方式调查了 30 个顾客购买商品的件数, 结果从小到大排列如下:

0	0	1	1	1	2	2	2	3	3
4	5	6	6	8	9	9	10	10	10
10	12	12	13	15	16	18	20	23	29

  求众数和样本均值.

- **解** 样本中 10 出现的次数最多, 是 4 次, 所以 10 是众数.
- 样本均值是

$$\bar{x} = \frac{1}{30}(1 + 1 + \cdots + 23 + 29) = 8.667.$$

- 在例 4.1 中, 如果购买件数最多的那个顾客购买件数增加从 29 增加的到 40, 众数不变, 样本均值增加为 9.04.
- 数据中最大值的变化对众数没有影响, 对样本均值的影响较大.

极差

- 数据中最大值与最小值的差称为**极差**, 直观反映了数据分布范围宽窄。
- 例 4.1 中极差为 $29 - 0 = 29$。

中位数

- 样本数据从小到大排列后，最中间的一个（有奇数个时）的值或最中间的两个（有偶数个时）的平均值为**样本中位数**。
- 记作 M_d。
- 用作样本左右位置的一个度量指标，以及总体左右位置的估计。
- 设观测数据已经被从小到大排列为:

$$x_1 \leq x_2 \leq \cdots \leq x_n.$$

- 对 n 为奇数情况，中位数是从小到大排在中间一个，$M_d = x_{\frac{n+1}{2}}$。如: $1, 5, 9, 12, 13$ 的中位数是 9.
- 对 n 为偶数情况，中位数用从小到大排在中间的两个平均计算, $M_d = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$。如: $1, 5, 9, 12, 13, 21$ 的中位数是 $M_d = \frac{9+12}{2} = 10.5$。
- 小于等于中位数的数据不少于样本量的二分之一, 大于等于中位数的数据不少于样本量的二分之一
- 中位数作为总体位置的度量比较不受极端值影响，而均值则会收到极端值影响。

例 4.2

- 2001 年, 在对 A 城市进行年人均收入调查时, 采用随机抽样的方法得到了以下 10 个数据 (单位: 万元)：

  0.79, 0.98, 1.17, 1.46, 1.67, 1.79, 1.82, 1.98, 2.26, 9.78

  计算中位数和样本均值.

- **解** 数据已经从小到大排列. 中位数是

$$M_d = \frac{1.67 + 1.79}{2} = 1.73.$$

- 样本均值

$$\bar{x} = \frac{0.97 + 0.98 + \cdots + 9.78}{10} = 2.37.$$

- 数据的茎叶图 (精确到小数点后 1 位):

茎	叶
0	8
1	02788
2	03
3	
4	
5	
6	
7	
8	
9	8

- 观测点 9.78 和其它值距离很远。称这样的值为**离群值**。
- 如果对 A 城市不了解其他经济情况, 很难作出年人均收入的合理估计.
- $\bar{x} = 2.37$ 很可能过高估计了总体均值, 因为 9.78 万元的年收入比其他人的年收入大的太多了.
- 9.78 拉高了样本均值.
- 中位数 1.73 看来好一些, 但是也可能低估或高估了总体均值, 我们还不了解有 9.78 万元年收入的人的比例.
- 要作出合理的估计, 必须增加抽样调查的样本量.

例 4.3

- 数学考试后, 甲班成绩的中位数是 72 分, 乙班成绩的中位数是 78 分. 仅从这两个数看, 哪班数学好的同学更多一些.
- **解** 甲班有不少于 50% 的同学的成绩在 72 分之下, 乙班有不少于 50% 的同学的学习成绩在 78 分之上, 乙班数学好的同学更多一些.

四分之一和四分之三分位数

- 设样本的中位数为 M_d，从样本中取出小于等于 M_d 的子集，计算这个子集的中位数，定义为四分之一分位数。

- 大于等于四分之一分位数的样本点占所有样本点比例约为四分之一，小于等于四分之一分位数的样本点占所有样本点比例约为四分之三。
- 四分之三分位数类似计算。

总体分位数

- 设总体 X 有严格单调增的连续分布函数 $F(x)$，$F^{-1}(p), p \in (0,1)$ 为 $F(\cdot)$ 的反函数。
- 对 $p \in (0,1)$，称 $F^{-1}(p)$ 为 X 的 p 分位数。
- 一般地，若 x_p 使得

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

  称 x_p 为 X 的 p 分位数。
- 这样的 p 分位数可能不唯一。
- 为了使得 p 分位数定义唯一，令

$$F^{-1}(p) = \inf\{x : F(x) \ge p\},\ p \in (0,1)$$

  称 $F^{-1}(p)$ 为 X 的 p 分位数。

样本分位数

- 样本分位数是总体分位数的估计。
- 对样本 $x_1, x_2, \ldots, x_n$，设其从小到大排列为 $x_{(1)} \le x_{(2)} \le \ldots x_{(n)}$，总体分位数 x_p 的估计常采用 $b = x_{([np])}$，这里 $[\cdot]$ 表示向下取整运算。
- 分位数估计有多种方法，$b = x_{([np])}$ 是一个最简单的估计。
- 约有比例 p 的样本小于等于样本 p 分位数，约有比例 $1-p$ 的样本大于等于样本 p 分位数。

6.5 随机对照试验

随机对照试验

- 观察研究是不严谨的研究方式。
- 为了解某药物、治疗等的效果，需要进行双盲随机对照试验。

- 例：坏血病研究。
- 例：经脉吻合分流试验。
- 随机对照组是必要的，否则可能得出错误结论。
- 例：脊髓灰质炎疫苗双盲随机对照试验。

例：坏血病的研究

- 17 世纪初期, 长期在海上航行的水手经常患坏血病. 坏血病的症状是牙龈肿大出血, 皮肤上出现青灰的斑点.
- 英国海军部试图考察坏血病的起因. 他们怀疑这是因为水手缺乏柑橘类的水果造成的.
- 当此想法提出时, 刚好有 4 艘军舰要远航. 为了调查是否由于水手缺乏柑橘类的水果而导致坏血病, 海军部随机地安排了一艘军舰上的水兵每天喝柑橘汁, 另外 3 艘军舰不供应柑橘汁. 这是一次安排好的试验.
- 试验的结果是: 航行还没有结束, 没有提供柑橘汁的水手多数得了坏血病, 而提供柑橘汁的军舰没有发现坏血病. 最后, 提供柑橘汁的军舰不得不把携带的柑橘汁分给其他的军舰, 以帮助他们顺利返航.
- 尽管本次试验的计划还可以从各个方面进行改进, 但是试验的结果成功地证实了最初的怀疑.
- 在例 5.1 中, 我们称喝柑橘汁的水兵是**试验组** (experimental group), 称不喝柑橘汁的水兵为**对照组** (control group).
- 试验组由随机选择出的对象构成, 试验组的成员要接受某种特殊的待遇或治疗等. 而对照组由那些没有接受这种特殊待遇的对象构成.
- 一个好的试验设计应当有一个试验组和一个对照组.
- 为什么要有对照组?
- 在例 5.1 中, 如果没有对照组, 为 4 艘军舰都提供柑橘汁, 就没有水兵患上坏血病, 海军部就不能确认他们的最初怀疑. 因为不能确定是否有其他的食品或治疗避免了坏血病.
- 为什么试验组要随机抽取?

- 在例 5.1 中, 如果安排喜欢喝柑橘汁的水兵在试验组, 喜欢喝啤酒的水兵在对照组, 就不能确定研究开始前这两组水兵的身体状况是否有差异. 水手身体状况的差异也可能影响是否容易得坏血病.
- 随机选择试验组能够有效地抵消个体差异造成的对试验结果的影响.
- 随机选择试验对象是英国统计学家费歇 (Fisher) 的贡献, 在 20 世纪初, 他用此方法致力于农业试验的研究. 从此随机选择试验组成为安排试验的基本原则.

例：静脉吻合分流术

- 在一些肝硬化病例中, 许多病人会从肝出血直至死亡. 历史上有一种称为“静脉吻合分流术”的外科手术用于治疗肝硬化, 其原理是运用外科手术的方法使血流改变方向.
- 这种手术花费很大并且有很高的危险性. 值得做这样的手术吗?
- 为了解决上述问题, 一共有三批共 51 次手术试验. 第一批进行了 32 次无对照组的试验. 结果如下:

设计方法	显著有效	中等有效	无效	试验次数
无对照组	24	7	1	32
所占比例	75%	21.9%	3.1%	100%

  试验说明有 75% 的手术显著有效, 21.9% 的手术中等有效, 看来手术是值得做的.

- 第二批共进行了 15 次手术试验, 这批试验有对照组, 但是对照组的病人不是随机选取的. 医生根据病人的临床诊断情况决定是将病人编入实验组做手术, 还是编入对照组不做手术. 结果如下:

设计方法	显著有效	中等有效	无效	试验次数
非随机对照	10	3	2	15
所占比例	66.7%	20%	13.3%	100%

  这次试验的结果是 66.7% 的手术显著有效, 20% 的手术中等有效, 13.3% 的手术无效. 这个试验结果也是对静脉吻合分流术的肯定. 这次的结果与无对照组的试验结果差别不是很大.

- 再看有随机选取的对照组的第三批试验, 这批试验只有 4 次手术. 随机选取的方式类似于掷硬币, 如果硬币正面朝上就将病人选入试验组

作手术. 这次试验处理组的结果如下:

设计方法	显著有效	中等有效	无效	试验次数
随机对照	0	1	3	4
所占比例	0%	25%	75%	100%

随机对照试验的结果显著地否定了外科手术“静脉吻合分流术”.

- 结果显示: 设计差的试验研究过分夸大了外科手术“静脉吻合分流术”的价值. 经过认真设计的试验研究显示“静脉吻合分流术”几乎没有什么价值.
- 为什么会出现如此大的差别呢?
- 在无对照组和非随机选取对照组的试验中, 实验者根据病人的临床诊断决定是否将他编入试验组进行手术.
- 这样做就出现一种自然的倾向: 试验人员更倾向于将那些身体状态较好的病人选入试验组, 以减少手术风险. 其结果有利于对手术的肯定评价, 这种结果是不真实的.
- 对上述试验的跟踪观测发现, 做手术的 51 个病人中 3 年后大约有 60% 仍然活着, 随机对照组中 (没做手术的病人)3 年后大约也有 60% 的病人仍然活者. 这就说明手术基本是无效的.
- 而在非随机对照组中, 只有 45% 的病人存活期超过三年, 这就说明了非随机对照组中的病人健康情况较差, 验证了健康情况较好的病人更容易被选入试验组作手术.
- **随机安排对照组是十分必要的, 否则可能得出错误的结论.**
- 我们称随机选取试验组的对照试验为**随机对照试验**.

其它随机对照试验案例

- 在人类历史上还有许多成功使用随机对照试验的例子, 也有许多惨重的教训.
- 例如, 随机对照实验否定了治疗冠状动脉病的冠状旁道外科手术 (该手术费用昂贵), 否定了用抗凝剂治疗心脏病突发, 否定了用 5-FU 对结肠癌进行化疗, 否定了用乙烯雌粉预防流产.

- 具体情况如下.

医疗方法	随机对照实验		非随机对照实验	
结论	有效	无效	有效	无效
冠状旁道手术	1	7	16	5
抗凝剂治疗	1	9	5	1
5-FU 结肠癌化疗	0	5	2	0
乙烯雌粉预防流产	0	3	5	0

- 特别需要指出的是有关乙烯雌粉的实验, 随机对照试验完全否定了这种预防流产的药. 但是起初的非随机对照试验却赞同药的疗效, 这是一个医学的悲剧. 在美国的 60 年代末, 医生每年大约为 5 万名孕妇发放这种药. 后来揭示, 怀孕期间的母亲服用乙烯雌粉, 20 年后给她们的女儿带来灾害性的副作用, 可能引发他们的女儿得一种罕见的癌症. 该药于 1971 年被禁止使用.

- 人们从太多的悲剧中总结了教训: 对一种新药不作随机对照实验是非常危险的.

小儿麻痹症疫苗随机对照双盲试验

- 1916 年小儿麻痹症 (脊髓灰质炎) 袭击了美国, 以后的 40 年间, 受害者成千上万. 20 世纪 50 年代, 人们开始开发预防疫苗. 当时萨凯 (Salk) 培育的疫苗最有希望. 他的疫苗在实验室中表现良好: 安全, 产生对脊髓灰质炎病毒的抗体. 但是在大规模使用前必须进行现场人体试验, 通过试验最后确定疫苗是否有效. 只有这样才能达到保护儿童的目的.

- 当时采用了随机对照的研究方案, 对每个儿童用类似投掷一个硬币的方法决定是否将其编入试验组: 正面朝上分在试验组, 否则分在对照组. 除了试验的设计人员, 连医生也不知道哪个儿童分在试验组, 哪个儿童分在对照组.

- 然后，给分在试验组的儿童注射疫苗, 给分在对照组的儿童注射生理盐水, 让他们认为也被注射了疫苗.

- 得到的结果如下:

	试验人数	试验后的发病率
试验组	20万	28/10万
对照组	20万	71/10万

- 试验结果显示, 疫苗将小儿麻痹症的发病率从 10 万分之 71 降低到 10 万分之 28. 由于 0.00071 和 0.00028 的差别超出了随机性本身所能解释的范围, 所以宣布疫苗是成功的.
- 进一步的分析指出, 可以以近 100% 的概率保证疫苗是有效的 (参考 §8.6, 例 6.4).
- 我们把对照组中的处理方法称为使用**安慰剂** (placebo), 例 5.3 中的安慰剂是注射生理盐水.
- 给对照组的儿童使用安慰剂是为了避免儿童的心理作用影响试验的结果. 尽管可以认为光靠精神作用不能抵抗小儿麻痹症, 但是为了确认试验结果的可靠性, 使用安慰剂是必要的.
- 不让医生知道儿童是来自试验组还是对照组是为了使医生能够作出更公正的诊断, 避免在诊断儿童是否患有小儿麻痹症时受到心理因素的影响.
- 此例中的随机对照试验又称为**随机对照双盲试验**. 双盲的之一是指儿童自己不知道自己是在试验组还是在试验组, 也就是说不知道自己被注射的是疫苗还是安慰剂, 甚至不知道有安慰剂, 这就有效地避免了潜在的心理影响.
- 另外一盲是指医生不了解他诊断的病人在对照组还是在试验组, 这就避免了医生对疫苗的主观看法带来的可能影响. 在可能的场合, 随机对照双盲试验可以最大程度地避免心理因素的影响.
- 在许多场合, 精神因素是不能忽视的. 有资料显示在医院中给那些手术后产生剧痛的病人服用由淀粉制成的“止痛片”后，大约有 1/3 的病人感觉剧痛减轻.

第七章　参数估计

7.1　点估计和矩估计

总体和样本

- 如果 X 是从总体中随机抽样得到的个体, 则 X 是随机变量, X 的分布就是总体的分布.
- 如果对总体进行有放回的随机抽样, 就得到独立同分布的, 和 X 同分布的随机变量 $X_1, X_2, \cdots, X_n$. 我们称 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的**简单随机样本**.
- 在 §2.2 的例 2.1, 观测放射性钋放射 α 粒子的试验中, 用 X 表示 7.5 秒内观测到的粒子数. 在独立重复观测时, 用 X_i 表示第 i 次观测结果, 则 $X_1, X_2, \cdots, X_n$ 独立同分布, 和 X 同分布, $X_1, X_2, \cdots, X_n$ 是来自总体 X 的简单随机样本.
- **定义 1.1** 如果 $X_1, X_2, \cdots, X_n$ 独立同分布, 和 X 同分布, 就称 X 是总体, 称 $X_1, X_2, \cdots, X_n$ 是总体 X 的简单随机样本, 称观测数据的个数 n 为样本量.
- 为了简单, 也把总体 X 的简单随机样本简称为总体 X 的样本.
- 在实际问题中得到的总是简单随机样本 $X_1, X_2, \cdots, X_n$ 的观测值 $x_1, x_2, \cdots, x_n$. 我们也称 $x_1, x_2, \cdots, x_n$ 是总体 X 的简单随机样本.
- 在统计学中, 常常不把 $X_1, X_2, \cdots, X_n$ 与它们的观测值 $x_1, x_2, \cdots, x_2$ 严格区分, 这是为了符号使用的方便.
- 当对数据进行统计分析时, 用大写的 $X_1, X_2, \cdots, X_n$, 实际计算时更多地用小写的 $x_1, x_2, \cdots, x_n$.
- 在统计问题中, 总体 X 的分布形式往往是已知的. 例如重复测量一个物体的重量时, 认为总体 X 服从正态分布 $N(\mu, \sigma^2)$, 未知参数是

(μ, σ^2), 问题是根据来自总体 X 的样本 $X_1, X_2, \cdots, X_n$ 估计总体参数 (μ, σ^2). 观测放射性钋放射 α 粒子时, 总体 X 服从泊松分布 $\mathrm{P}(\lambda)$, 未知参数是 λ, 问题是根据来自总体 X 的样本 $X_1, X_2, \cdots, X_n$ 估计 λ.

估计量 (统计量)

- 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的简单随机样本, θ 是总体 X 的未知参数. 如果 $g(x_1, x_2, \cdots, x_n)$ 是已知函数, 就称

$$\hat{\theta} = g(X_1, X_2, \cdots, X_n)$$

是 θ 的**估计量**, 简称为**估计** (estimator). 换句话说, 估计或估计量是从观测数据 $X_1, X_2, \cdots, X_n$ 能够直接计算的量. 计算后得到的值称为**估计值**. 估计量也称为**统计量** (statistic).

- 设 $\hat{\theta}$ 是总体参数 θ 的估计, 作为随机变量 $X_1, X_2, \cdots, X_n$ 的函数, 估计量 $\hat{\theta}$ 也是随机变量. 估计量是样本的函数.

估计的优良性

- 用估计量 $\hat{\theta}$ 去估计总体参数 θ，我们希望 $\hat{\theta}$ 能够尽可能与 θ 接近，但由于随机性影响误差是不可避免的。

- 引入下面的关于估计优良性的定义。

- **定义 1.2** 设 $\hat{\theta}$ 是 θ 的估计.

  (1) 如果 $\mathrm{E}\hat{\theta} = \theta$, 称 $\hat{\theta}$ 是 θ 的**无偏估计**;

  (2) 如果当样本量 $n \to \infty$, $\hat{\theta}$ 依概率收敛到 θ, 就称 $\hat{\theta}$ 是 θ 的**相合估计** (consistent estimator);

  (3) 如果当样本量 $n \to \infty$, $\hat{\theta}$ 以概率 1 收敛到 θ, 就称 $\hat{\theta}$ 是 θ 的**强相合估计** (strongly consistent estimator).

- 由于以概率 1 收敛可以推出依概率收敛, 所以强相合估计一定是相合估计.

- 一个估计起码应当是相合的, 否则我们不知道这个估计有什么优点, 也不知道它估计的是谁.

均值的估计

- 设总体均值 $\mu = \mathrm{E}X$ 存在, $X_1, X_2, \cdots, X_n$ 是总体 X 的简单随机样本.

- 均值 μ 的估计定义为

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.1}$$

- 由于 $\bar{X}_n$ 是从样本计算出来的, 所以是样本均值.

- 样本均值 $\bar{X}_n$ 有如下的性质.

  (1) $\bar{X}_n$ 是 μ 的无偏估计. 这是因为 $\mathrm{E}\bar{X}_n = \mu$.

  (2) $\bar{X}_n$ 是 μ 的强相合估计, 从而是相合估计. 这是因为从强大数律得到

$$\lim_{n\to\infty} \bar{X}_n = \mu, \mathrm{wp1}. \tag{1.2}$$

方差的估计

- 总体方差 $\sigma^2 = \mathrm{Var}(X)$ 的点估计由

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \hat{\mu})^2 \tag{1.4}$$

  定义. 由于 S^2 是从样本计算出来的, 所以是样本方差.

- 定义 $Y_j = X_j - \mu$, 有

$$\begin{aligned}
&\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j = \hat{\mu} - \mu,\\
&Y_j - \bar{Y}_n = X_j - \hat{\mu},\\
&\mathrm{E}\bar{Y}_n^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

- 于是得到

$$\begin{aligned} S^2 =& \frac{1}{n-1}\sum_{j=1}^{n}(X_j-\bar{X}_n)^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j-\bar{Y}_n)^2 \\ =& \frac{1}{n-1}\sum_{j=1}^{n}(Y_j^2-2Y_j\bar{Y}_n+\bar{Y}_n^2) \\ =& \frac{1}{n-1}\Big[\sum_{j=1}^{n}Y_j^2-2n\bar{Y}_n\bar{Y}_n+n\bar{Y}_n^2\Big] \\ =& \frac{1}{n-1}\Big[\sum_{j=1}^{n}Y_j^2-n\bar{Y}_n^2\Big]. \end{aligned} \tag{1.5}$$

- 从而有

$$\mathrm{E}S^2 = \frac{1}{n-1}\Big[\sum_{j=1}^{n}\mathrm{E}Y_j^2-n\mathrm{E}\bar{Y}_n^2\Big] = \frac{1}{n-1}(n\sigma^2-\sigma^2) = \sigma^2.$$

  说明 S^2 是 σ^2 的无偏估计.

- 利用强大数律得到

$$\frac{1}{n-1}\sum_{j=1}^{n}Y_j^2 \to \mathrm{E}Y_1^2 = \sigma^2, \mathrm{wp1}.$$

$$\frac{1}{n-1}n\bar{Y}_n^2 \to (\mathrm{E}Y_1)^2 = 0, \mathrm{wp1}.$$

  所以由 1.5 得到

$$S^2 \to \sigma^2, \mathrm{wp1}. \tag{1.6}$$

  说明 S^2 是强相合估计, 从而也是相合估计.

标准差 σ 的估计

- 由于 S^2 是 σ^2 的估计, 所以定义标准差 σ 的估计为

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j-\hat{\mu})^2}.$$

- S 是样本标准差. 由于 $S \to \sigma, \mathrm{wp1}.$ 成立, 所以 S 是 σ 的强相合估计.

- 但是 S 一般不是 σ 的无偏估计. 实际上用内积不等式得到

$$\mathrm{E}S = \mathrm{E}(1\cdot S) \le \sqrt{1\cdot \mathrm{E}S^2} = \sigma.$$

  等号成立时有不全为零的常数 a, b 使得 $P(aS+b=0)=1$, 于是 $S = b/a$, wp1. 所以只要 S 等于常数的概率小于 1, 则 $\mathrm{E}S < \sigma$.

样本均值、方差、标准差的理论结果

- **定理 1.1** 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的简单随机样本, $\mu = \mathrm{E}X$, $\sigma^2 = \mathrm{Var}(X)$.

  (1) 样本均值 $\bar{X}_n$ 是总体均值 μ 的强相合无偏估计,

  (2) 样本方差 S^2 是总体方差 σ^2 的强相合无偏估计,

  (3) 样本标准差 S 是总体标准差 σ 的强相合估计.

例 1.1

- 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的简单随机样本, 则 $X_1^j, X_2^j, \cdots, X_n^j$ 是总体 X^j 的简单随机样本, 所以当原点矩 $\nu_j = \mathrm{E}X^j$ 存在时,

$$\hat{\nu}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j \tag{1.7}$$

  是 ν_j 的点估计.

- $\hat{\nu}_j$ 具有无偏性和强相合性.

- 最后指出, 在实际数据的计算中, 也常用 $\bar{x}_n$, s^2 和 s 分别表示样本均值, 样本方差和样本标准差:

$$\bar{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j,\ s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x}_n)^2,\ s = \sqrt{s^2}. \tag{1.8}$$

矩估计

- 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的简单随机样本, 已知 X 有分布函数

$$F(x; \theta_1, \theta_2, \cdots, \theta_m). \tag{1.9}$$

  其中的 $\theta_1, \theta_2, \cdots, \theta_m$ 是未知参数.

- 如果能得到表达式

$$\begin{cases} \theta_1 = g_1(\nu_1, \nu_2, \cdots, \nu_m), \\ \theta_2 = g_2(\nu_1, \nu_2, \cdots, \nu_m), \\ \cdots\cdots\cdots\cdots\cdots, \\ \theta_m = g_m(\nu_1, \nu_2, \cdots, \nu_m), \end{cases} \tag{1.10}$$

  其中

$$\nu_j = \mathrm{E}X^j,\ j = 1, 2, \cdots, m,$$

- 就称由

$$\begin{cases} \hat{\theta}_1 = g_1(\hat{\nu}_1, \hat{\nu}_2, \cdots, \hat{\nu}_m), \\ \hat{\theta}_2 = g_2(\hat{\nu}_1, \hat{\nu}_2, \cdots, \hat{\nu}_m), \\ \cdots\cdots\cdots\cdots\cdots, \\ \hat{\theta}_m = g_m(\hat{\nu}_1, \hat{\nu}_2, \cdots, \hat{\nu}_m) \end{cases} \tag{1.11}$$

  定义的 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$ 分别是 $\theta_1, \theta_2, \cdots, \theta_m$ 的**矩估计** (moment estimator). 这里的 $\hat{\nu}_j$ 是 ν_j 的点估计, 由 (1.7) 定义.

- 由于总体分布 (1.9) 中含有未知参数, 所以 ν_j 是参数 $\theta_1, \theta_2, \cdots, \theta_m$ 的函数, 而方程 (1.10) 通常是由下面的估计方程

$$\begin{cases} \nu_1 = h_1(\theta_1, \theta_2, \cdots, \theta_m), \\ \nu_2 = h_2(\theta_1, \theta_2, \cdots, \theta_m), \\ \cdots\cdots\cdots\cdots\cdots, \\ \nu_m = h_m(\theta_1, \theta_2, \cdots, \theta_m) \end{cases} \tag{1.12}$$

  得到的. 注意这里的 $\nu_j = \mathrm{E}X^j$.

正态分布参数的矩估计

- 设 X 服从正态分布 $N(\mu, \sigma^2)$.
- 由于

$$\mu = \mathrm{E}X,\ \sigma^2 = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \nu_2 - \nu_1^2,$$

- 所以 μ, σ^2 的矩估计分别是

$$\begin{aligned} \hat{\mu} =& \bar{X}_n, \\ \hat{\sigma}^2 =& \hat{\nu}_2 - (\hat{\nu}_1)^2 \\ =& \frac{1}{n}\sum_{j=1}^{n} X_j^2 - (\bar{X}_n)^2 \\ =& \frac{1}{n}\sum_{j=1}^{n} (X_j - \hat{\mu})^2. \end{aligned}$$

指数分布参数的矩估计

- 设 X 服从指数分布 $\mathrm{E}(\lambda)$.

- 由 $\mu = \mathrm{E}X = 1/\lambda$ 得到 $\lambda = 1/\mu$,
- 从而得到矩估计

$$\hat{\lambda} = 1/\bar{X}_n.$$

泊松分布参数的矩估计

- 设 X 服从泊松分布 $\mathrm{P}(\lambda)$.
- 由 $\mu = \mathrm{E}X = \lambda$,
- 得到矩估计 $\hat{\lambda} = \bar{X}_n$.

均匀分布参数的矩估计

- X 服从均匀分布 $\mathrm{U}(0, b)$.
- 由 $\mu = b/2$,
- 得到 $b = 2\mu$,
- 从而得到矩估计 $\hat{b} = 2\bar{X}_n$.

二项分布参数的矩估计

- X 服从二项分布 $B(m, p)$, 其中 m 已知.
- 由 $\mu = \mathrm{E}X = mp$ 得到 $p = \mu/m$,
- 从而得到 p 的矩估计 $\hat{p} = \bar{X}_n/m$.

几个矩估计的性质

- 以上的几个矩估计中, 除了正态分布的 $\hat{\sigma}^2$ 和指数分布中的 $\hat{\lambda}$ 不是无偏估计外, 其余的都是无偏估计.
- 以上五个矩估计都是强相合估计.

例 1.2

- 设一大批产品的合格率是 p, 每次从中随机抽出 10 件进行检验.
- 用 X_i 表示第 i 次抽出的 10 件中次品的个数, 则可以认为 $X_1, X_2, \cdots, X_n$ 独立同分布, 总体分布是二项分布 $B(10, p)$.
- 用 $\mu = \mathrm{E}X$ 表示 $B(10, p)$ 的数学期望, 由 $\mathrm{E}X = 10p$, 得到 $p = \frac{\mathrm{E}X}{10}$.
- 于是 p 的矩估计是
$$\hat{p} = \frac{1}{10}\bar{X}_n$$
- 给定 n 次抽样的观测数据 $x_1, x_2, \cdots, x_n$ 后, p 的矩估计
$$\hat{p} = \frac{x_1 + x_2 + \cdots + x_n}{10n}.$$

7.2 最大似然估计

7.2.1 离散型随机变量的情况

例 2.1

- 某试验成功的概率是 $p = 0.9$ 或 $p = 0.1$. 现在试验成功了, 应当判断 $p = 0.9$ 还是判断 $p = 0.1$?
- 答案是明显的, 应当判断 $p = 0.9$.
- 例 2.1 提示我们, 试验能成功是因为成功的概率较大.
- 更一般地说, 我们能够观测到一个事件是因为这个事件发生的概率较大.
- 这样思考问题的思想被称为最大似然思想.

例 2.2

- 设 $X_1, X_2, \cdots, X_n$ 独立同分布, 都服从 Poisson 分布 Poisson(λ).

  (1) 给定观测 $X_1 = x_1$, 试估计 λ,

  (2) 给定观测数据 $x_1, x_2, \cdots, x_n$, 试估计 λ.

- **解** (1) X_1 有概率分布
$$P(X_1 = x) = \frac{\lambda^x}{x!}e^{-\lambda},\ x = 0, 1, \cdots.$$

- 按照最大似然估计思想, 观测到 $X_1 = x_1$ 的原因是观测到 x_1 的概率较大.
- 于是参数 λ 应当使得

$$L(\lambda) = \frac{\lambda^{x_1}}{x_1!}e^{-\lambda}, \lambda > 0,$$

达到最大值.

- 于是问题简化为求 $L(\lambda)$ 的最大值的问题. 为了计算方便, 引入

$$l(\lambda) = \ln L(\lambda) = x_1 \ln \lambda - \lambda - \ln(x_1!),$$

由 $l'(\lambda) = x_1/\lambda - 1 = 0$, 知道应当取 $\lambda = x_1$.

- (2) 根据最大似然估计的思想, 参数 λ 应当使得观测到 $(x_1, x_2, \cdots, x_n)$ 的概率最大.
- 这等价于 λ 使得 $(X_1, X_2, \cdots, X_n)$ 的联合分布

$$\begin{aligned} L(\lambda) =& P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) \\ =& \frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}e^{-\lambda}\cdots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda} \\ =& \frac{\lambda^{(x_1+x_2+\cdots+x_n)}}{x_1!x_2!\cdots x_n!}e^{-n\lambda}. \end{aligned}$$

取得最大值.

- 于是应当用 $L(\lambda)$ 的最大值点

$$\hat{\lambda} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

作为 λ 的估计.

- 注意在上述问题中, $(x_1, x_2, \cdots, x_n)$ 是已经观测到的数据, 所以 $L(\lambda)$ 是 λ 的函数, 称为基于数据 $(x_1, x_2, \cdots, x_n)$ 的似然函数, 简称为**似然函数** (likelihood function).

最大似然估计定义 (离散情况)

- **定义 2.1** 设离散随机变量 $X_1, X_2, \cdots, X_n$ 有联合分布

$$p(x_1, x_2, \cdots, x_n; \theta) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n),$$

其中 θ 是未知参数, 给定观测数据 $x_1, x_2, \cdots, x_n$ 后, 我们称 θ 的函数

$$L(\theta) = p(x_1, x_2, \cdots, x_n; \theta)$$

为基于 $x_1, x_2, \cdots, x_n$ 的似然函数, 称 $L(\theta)$ 的最大值点 $\hat{\theta}$ 为 θ 的**最大似然估计** (maximum likelihood estimator).

- 定义 2.1 中的 θ 也可以是向量 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$.

7.2.2 连续型随机变量的情况

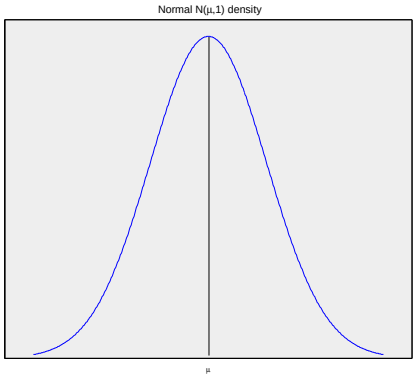
例 2.3

- 设 $X \sim N(\mu, 1)$, 给定观测数据 $X = x$, 如何估计 μ 呢?
- 根据正态密度函数的图形和最大似然思想, 知道 μ 应当在 x 的附近.
- 如果取 μ 大于 x, 就会问为什么不取 μ 小于 x.
- 因而最好取 $\mu = x$.
- 注意 X 有密度函数

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

给定观测 $X = x$, $\mu = x$ 恰好使得 $f(x;\mu)$ 达到最大.

一元正态分布 N(μ,1) 的密度函数



例 2.4

- 设 X_1, X_2 独立同分布, 都服从正态分布 $N(\mu, \sigma^2)$. 给定观测数据 x_1, x_2 如何估计 μ, σ^2 呢?
- 我们知道 (X_1, X_2) 的联合密度

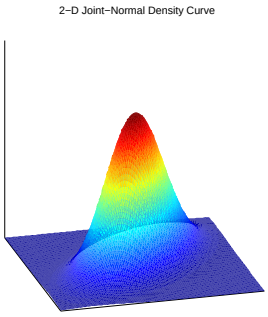
$$f(x_1, x_2; \mu, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\Big(- \frac{1}{2\sigma^2}[(x_1 - \mu)^2 + (x_2 - \mu)^2]\Big)$$

  是一个扣在 x, y 平面上的单峰曲面.
- 根据最大似然思想, μ, σ^2 的选择应当使得曲面在 (x_1, x_2) 处达到最大.
- 于是使得函数

$$L(\mu, \sigma^2) = f(x_1, x_2; \mu, \sigma^2)$$

  达到最大值的 $(\hat{\mu}, \hat{\sigma}^2)$ 就是参数 (μ, σ^2) 的估计.
- 于是问题转化为求 $L(\mu, \sigma^2)$ 的最大值点的问题.
- 注意上面的 x_1, x_2 已经是常数了, 自变元是 (μ, σ^2).

二元正态分布的密度函数曲面



最大似然估计 (连续型)

- **定义 2.2** 设随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 有联合密度 $f(\boldsymbol{x}; \boldsymbol{\theta})$, 其中 $\boldsymbol{\theta}$ 是未知参数. 得到 $\boldsymbol{X}$ 的观测值 $\boldsymbol{x}$ 后, 称 $\boldsymbol{\theta}$ 的函数

$$L(\boldsymbol{\theta}) = f(\boldsymbol{x}; \boldsymbol{\theta})$$

  为基于 $\boldsymbol{x}$ 的**似然函数**. 称似然函数 $L(\boldsymbol{\theta})$ 的最大值点 $\hat{\boldsymbol{\theta}}$ 为参数 $\boldsymbol{\theta}$ 的**最大似然估计**.

- 最大似然估计通常被缩写成 **MLE**(Maximum Likelihood Estimator).

- 设总体 X 有密度函数 $f(x; \boldsymbol{\theta})$, $X_1, X_2, \cdots, X_n$ 是总体 X 的简单随机样本, 则 $(X_1, X_2, \cdots, X_n)$ 的联合密度是

$$f(x_1, x_2, \cdots, x_n; \boldsymbol{\theta}) = \prod_{j=1}^{n} f(x_j; \boldsymbol{\theta}),$$

- 基于观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 的似然函数是

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n} f(x_j; \boldsymbol{\theta}). \tag{2.1}$$

- 由于

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) \tag{2.2}$$

  和似然函数 (2.1) 有相同的最大值点, 所以称 (2.2) 为**对数似然函数**. 实际问题中, 求对数似然函数 $l(\boldsymbol{\theta})$ 的最大值点往往要方便得多.

例: 正态总体参数的 MLE

- 设总体 X 服从正态分布 $N(\mu, \sigma^2)$.

- 由于总体 X 的密度函数是

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi a}} \exp\big[-\frac{(x-\mu)^2}{2a}\big], \text{ 其中 } a \triangleq \sigma^2.$$

- 所以基于观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 的似然函数是

$$L(\mu, a) = \frac{1}{(\sqrt{2\pi a})^n} \exp\big[-\sum_{j=1}^{n} \frac{(x_j-\mu)^2}{2a}\big].$$

- 对数似然函数是

$$l(\mu, a) = -\frac{n}{2}\ln a - \sum_{j=1}^{n}\frac{(x_j-\mu)^2}{2a} - \frac{n}{2}\ln(2\pi).$$

- 求 $l(\mu, a)$ 的最大值点可以通过解方程组

$$\begin{cases} \frac{\partial l}{\partial \mu} = \frac{1}{a}\sum_{j=1}^{n}(x_j-\mu) = 0, \\ \frac{\partial l}{\partial a} = -\frac{n}{2a} + \frac{1}{2a^2}\sum_{j=1}^{n}(x_j-\mu)^2 = 0 \end{cases} \tag{2.3}$$

得到.

- 从 (2.3) 解得 μ, $\sigma^2 = a$ 的 MLE 为

$$\begin{cases} \hat{\mu} = \bar{X}_n; \\ \hat{\sigma^2} = \hat{a} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{X}_n)^2. \end{cases}$$

- 可以看出, 对于正态总体来讲, 最大似然估计和矩估计是一致的.

指数分布总体参数的 MLE

- 设总体 X 服从指数分布 E(λ).

- 由于指数分布 E(λ) 的密度函数是

$$f(x;\lambda) = \lambda e^{-\lambda x},\ x \geq 0.$$

- 基于观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 的似然函数是

$$L(\lambda) = \lambda^n \exp\big(-\lambda\sum_{j=1}^{n}x_j\big)$$

- 对数似然函数是

$$l(\lambda) = n\ln\lambda - \lambda\sum_{j=1}^{n}x_j.$$

- 由

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{j=1}^{n}x_j = 0,$$

得到参数 λ 的 MLE 为 $\hat{\lambda} = 1/\bar{X}_n$.

均匀分布总体参数的 MLE

- 设总体 X 服从均匀分布 $\mathrm{U}[0,b]$.
- 密度函数是
$$f(x;b)=\frac{1}{b}\mathrm{I}_{\{0\le x\le b\}}.$$
- 给定观测数据 $x_1,x_2,\cdots,x_n$, 定义
$$\begin{aligned} x_{(1)}&=\min\{x_1,x_2,\cdots,x_n\},\\ x_{(n)}&=\max\{x_1,x_2,\cdots,x_n\}. \end{aligned}\tag{2.4}$$
- 可以把似然函数写成
$$L(b)=\frac{1}{b^n}\prod_{j=1}^{n}\mathrm{I}_{\{0\le x_i\le b\}}=\frac{1}{b^n}\mathrm{I}_{\{0\le x_{(1)}\le x_{(n)}\le b\}}.$$
- 要 $L(b)$ 达到最大, 首先要示性函数 $\mathrm{I}_{\{0\le x_{(1)}\le x_{(n)}\le b\}}=1$, 即要 $b\ge x_{(n)}$.
- 然后再要求 $1/b^n$ 尽量的大.
- 不难看出, 这时必须取 $b=x_{(n)}$, 所以 b 的 MLE 是 $\hat{b}=x_{(n)}$.

均匀分布最大似然估计和矩估计的比较

- 对于均匀分布来将, 矩估计是 $\tilde{b}=2\bar{x}_n$, 最大似然估计是 $\hat{b}=x_{(n)}$. 这两者明显是不一样的.
- 为了解那个估计更准确一些, 我们用计算机产生一万个在区间 $[0,2.8]$ 上均匀分布的**随机数** (独立同分布随机变量的观测值), 使用前 n 个随机数计算出矩估计 $\tilde{b}$ 和 MLE $\hat{b}$ 的估计误差列在下面的表中.
- $b=2.8$

$n=$	10	30	60	100	1000	10000
$\lvert\tilde{b}-b\rvert=$	0.231	0.315	0.313	0.063	0.096	0.003
$\lvert\hat{b}-b\rvert=$	0.249	0.249	0.214	0.008	0.004	0.00002

- 我们称上述的方法为计算机模拟试验方法. 从上述模拟试验看出, MLE $\hat{b}$ 的表现似乎要比矩估计 $\tilde{b}$ 好.
- 但是数据的产生带有随机性, 所以一次模拟试验显然不够.
- 为了克服数据的随机性, 我们将上述模拟试验独立重复 1000 次.

- 用 $\tilde{b}_j$ 表示第 j 次模拟计算得到的矩估计 $\tilde{b}$, 用 $\hat{b}_j$ 表示第 j 次模拟计算得到的最大似然估计 $\hat{b}$.

- 再定义 $m = 1000$ 次模拟的平均

$$M(\tilde{b}) = \frac{1}{m}\sum_{j=1}^{m}\tilde{b}_j, \qquad M(\hat{b}) = \frac{1}{m}\sum_{j=1}^{m}\hat{b}_j.$$

- 和 $m = 1000$ 次模拟的样本标准差

$$std(\tilde{b}) = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}(\tilde{b}_j - \tilde{b})^2},$$
$$std(\hat{b}) = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}(\hat{b}_j - \bar{b})^2}.$$

- 结论列入下表.

$n =$	10	30	60	100	1000	10000
$\lvert M(\tilde{b}) - b\rvert =$	0.007	0.014	0.011	0.008	0.002	0.0006
$std(\tilde{b}) =$	0.511	0.293	0.203	0.162	0.048	0.016
$\lvert M(\hat{b}) - b\rvert =$	0.270	0.096	0.048	0.029	0.003	0.0002
$std(\hat{b}) =$	0.2398	0.0989	0.0466	0.0281	0.0028	0.0003

- 从上述计算中看出, $|M(\tilde{b}) - b|$ 普遍小于 $|M(\hat{b}) - b|$, 所以从偏差的角度讲, 矩估计好一些.

- 这个结果是和 $\mathrm{E}\bar{X}_n = b$, $\mathrm{E}\hat{b} < b$ 一致的.

- 但是另一方面, $std(\hat{b})$ 普遍小于 $std(\tilde{b})$, 说明最大似然估计的稳定性更好一些.

- 由于一千次重复试验已经能够较好的克服随机因素的影响, 所以可以认为上述模拟试验的结果是可信的.

7.3 抽样分布及其上 α 分位数

抽样分布及其上 α 分位数—引言

- 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的简单随机样本, 当 $X \sim N(\mu, \sigma^2)$ 时, 也称 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的简单随机样本.

- 仍用

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j,$$
$$S^2 = \frac{1}{(n-1)}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2 \tag{3.1}$$

  分别表示样本均值和样本方差.

- 为了进一步研究未知参数的统计推断问题, 本节介绍几个重要的抽样分布. 我们将在以后的章节中使用这里介绍的分布.

- 所谓抽样分布, 意指基于样本构造的统计量的概率分布.

- 以后把来自总体 X 的简单随机样本简称为来自总体 X 的样本.

7.3.1 抽样分布

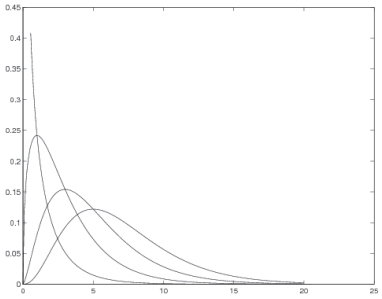
卡方分布

- **定义 3.1** (χ^2 分布 (卡方分布)) 如果随机变量 ξ 有概率密度

$$p(u) = \frac{1}{2^{n/2}\Gamma(n/2)} u^{\frac{n}{2}-1} e^{-u/2}, \quad u \geq 0. \tag{3.2}$$

  就称 ξ 服从 n 个自由度的 χ^2 分布, 记做 $\xi \sim \chi^2(n)$.

- 卡方 $\chi^2(n)$ 分布密度的图形, 按纵坐标的最大值从高到低自由度依次是 $n = 1, 3, 5, 7$:



χ^2 分布与正态分布的关系

- **定理 3.1** 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(0,1)$ 的样本, 则平方和

$$\xi_n = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n). \tag{3.3}$$

- 用归纳法可以证明本定理.
- **定理 3.2** 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(0,1)$ 的样本, 则有如下的结果.

  (1) $\overline{X}_n$ 和 S^2 独立;

  (2) $(n-1)S^2 \sim \chi^2(n-1)$.
- **推论 3.3** 如果 $\xi \sim \chi^2(n)$, $\eta \sim \chi^2(m)$, 则

  (1) 则 $\mathrm{E}\xi = n$,

  (2) 当 ξ 和 η 独立, 有 $\xi + \eta \sim \chi^2(n+m)$.
- **证** (1) ξ 和 (3.3) 中的 ξ_n 同分布, 所以有相同的数学期望 n.
- (2) 设 $X_1, X_2, \cdots, X_{n+m}$ 是来自总体 $N(0,1)$ 的样本, 则 $\xi + \eta$ 和

  $$\xi_n + \eta_m \stackrel{\triangle}{=} (X_1^2 + X_2^2 + \cdots + X_n^2) + (X_{n+1}^2 + X_{n+2}^2 + \cdots + X_{n+m}^2)$$

  同分布, 所以结论成立.
- **定理 3.4** 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, 则

  (1) $\overline{X}_n$ 和 S^2 独立;

  (2) $\frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2 \sim \chi^2(n-1)$.
- **证**: 设 $Y_j = (X_j - \mu)/\sigma$, 则 $Y_1, Y_2, \cdots, Y_n$ 是来自总体 $N(0,1)$ 的样本, 且

$$\begin{aligned}\bar{Y}_n =& \frac{1}{\sigma}(\bar{X}_n - \mu),\\ (n-1)S_Y^2 =& \sum_{j=1}^{n}(Y_j - \overline{Y}_n)^2 = \frac{1}{\sigma^2}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2\\ =& \frac{(n-1)}{\sigma^2}S^2.\end{aligned}$$

- 根据定理 3.2, $\bar{Y}_n$ 和 S_Y^2 独立, 从而知道 $\overline{X}_n = \sigma\bar{Y}_n + \mu$ 和 $S^2 = \sigma^2 S_Y^2$ 独立.
- 又由定理 3.2 知道

$$\frac{(n-1)}{\sigma^2}S^2 = (n-1)S_Y^2 \sim \chi^2(n-1).$$

t 分布

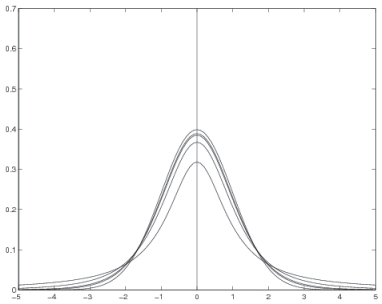
- **定义 3.2 (t 分布)** 如果随机变量 T 有概率密度

$$p_n(u) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\Big(1+\frac{u^2}{n}\Big)^{-\frac{n+1}{2}},\ u \in (-\infty, \infty).$$

  就称 T 服从 n 个自由度的 t 分布, 记做 $T \sim t(n)$.

t 分布分布密度

- t 分布的密度函数比正态分布的密度函数低一点.
- 下图是 $t(n)$ 分布密度函数的图形, 按纵坐标的最大值从低到高, 自由度依次是 $n = 1, 3, 7, 10$. 最高的一条曲线是标准正态密度曲线.



- 从中看出 n 增大时, $t(n)$ 分布向 $N(0,1)$ 分布的收敛是很快的. 当 $n \geq 33$, $t(n)$ 分布的密度和 $N(0,1)$ 的密度几乎就没有差别了.
- 实际上可以验证当 $n \geq 33$, 对标准正态密度函数 $\varphi(x)$ 有

$$\sup_x |p_n(x) - \varphi(x)| \leq 0.0041.$$

t 分布与正态分布及卡方分布的关系

- **定理 3.5** 如果 $Z \sim N(0,1)$, $\eta \sim \chi^2(n)$, Z, η 独立, 则

$$\frac{Z}{\sqrt{\eta/n}} \sim t(n).$$

- **定理 3.6** 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, 则

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S} \sim t(n-1).$$

- **证** 由定理 3.4 知道

$$Z=\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\sim N(0,1),\ \xi=\frac{n-1}{\sigma^2}S^2\sim\chi^2(n-1),$$

Z 和 ξ 独立, 于是用定理 3.5 得到

$$\frac{\bar{X}_n-\mu}{S/\sqrt{n}}=\frac{Z}{\sqrt{\xi/(n-1)}}\sim t(n-1).$$

F 分布

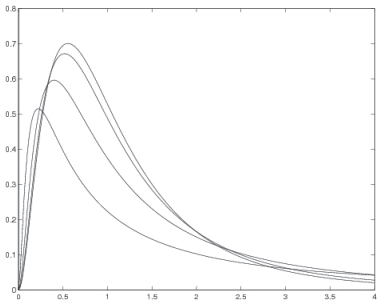
- **定义 3.3 ($F(n,m)$ 分布)** 如果随机变量 F 有概率密度

$$p(u)=\frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\Big(\frac{n}{m}\Big)^{\frac{n}{2}}\Big(1+\frac{n}{m}u\Big)^{-\frac{n+m}{2}}u^{\frac{n}{2}-1},\ u\geq 0,$$

就称 F 服从自由度为 (n,m) 的 F 分布, 记做 $F\sim F(n,m)$.

- 我们又称 n 是第一自由度, m 是第二自由度.

- 下图是 $F(6,m)$ 的分布密度函数图形, 按纵坐标的最大值从小到大, 第二自由度依次是 $m=1,3,7,10$.



F 分布与卡方分布的关系

- **定理 3.7** 如果 $\xi\sim\chi^2(n)$, $\eta\sim\chi^2(m)$, ξ 和 η 独立, 则

$$F=\frac{\xi/n}{\eta/m}=\frac{m\xi}{n\eta}\sim F(n,m),$$

$$F^{-1}=\frac{\eta/m}{\xi/n}=\frac{n\eta}{m\xi}\sim F(m,n).$$

F 分布与正态分布的关系

- **定理 3.8** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, $Y_1, Y_2, \cdots, Y_m$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, 又设这两个总体是相互独立的. 则当 $n, m \geq 2$,

$$S_X^2 / S_Y^2 \sim F(n-1, m-1).$$

其中

$$S_X^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2, \;\; S_Y^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \bar{Y}_m)^2.$$

- **证明:**

- 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, $Y_1, Y_2, \cdots, Y_m$ 是来自总体 Y 的样本.

- 如果总体 X 和总体 Y 独立, 则来自这两个总体的样本也是相互独立的, 于是

$$X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_m$$

是相互独立的随机变量.

- 所以 S_X^2 与 S_Y^2 独立。

- 由定理 3.4 知道

$$\xi = \frac{n-1}{\sigma^2} S_X^2 \sim \chi^2(n-1), \; \eta = \frac{m-1}{\sigma^2} S_Y^2 \sim \chi^2(m-1),$$

而 ξ, η 又是独立的, 所以用定理 3.7 得到

$$\frac{S_X^2}{S_Y^2} = \frac{\xi/(n-1)}{\eta/(m-1)} \sim F(n-1, m-1).$$

依分布收敛的一个定理

- 下面的定理也是后面常用的结论. 回忆 $\Phi(x)$ 总表示标准正态分布的分布函数, $\xi_n \to^d \xi$ 表示 ξ_n 依分布收敛到 ξ.

- **定理 3.9** 设 ξ_n 依分布收敛到 $N(0,1)$:

$$P(\xi_n \leq x) \to \Phi(x)$$

. 如果 $\eta_n \to 1, \text{wp1.}$, 则 $\xi_n \eta_n$ 也依分布收敛到 $N(0,1)$.

例 3.1

- 设 $\mu = \mathrm{E}X$, $\sigma^2 = \mathrm{Var}(X)$ 是正数, $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, 则
$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \to^d N(0,1).$$
- **证** 由中心极限定理知道
$$\xi_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \to^d N(0,1).$$
- 样本标准差 S 是 σ 的强相合估计: $S \to \sigma$, wp1.,
- 于是
$$\eta_n = \sigma/S \to 1, \text{ wp1.}$$
- 最后利用定理 3.9 得到
$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \frac{\sigma}{S} = \xi_n \eta_n \to^d N(0,1).$$
- 在例 3.1 中, 如果总体 $X \sim N(\mu, \sigma^2)$, 则
$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1).$$
- 对较大的 n, 又有
$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim N(0,1)$$
近似成立.
- 所以当 n 较大, $t(n)$ 分布和标准正态分布相近.

7.3.2 抽样分布的上 α 分位数

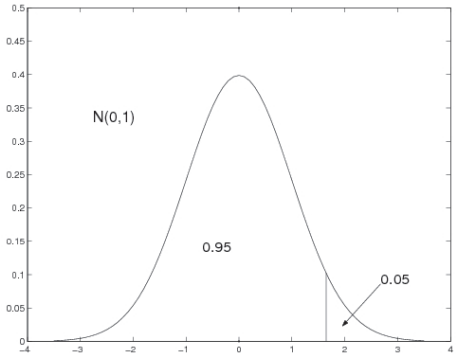
抽样分布的上 α 分位数

- 设正数 $\alpha \in (0,1)$.
- 对 $Z \sim N(0,1)$, 有唯一的 z_α 使得 $P(Z \geq z_\alpha) = \alpha$,
- 对 $\xi_n \sim \chi^2(n)$, 有唯一的 $\chi^2_\alpha(n)$ 使得 $P\big(\xi_n \geq \chi^2_\alpha(n)\big) = \alpha$,
- 对 $T_n \sim t(n)$, 有唯一的 $t_\alpha(n)$ 使得 $P(T_n \geq t_\alpha(n)) = \alpha$,
- 对 $F_{n,m} \sim F(n,m)$, 有唯一的 $F_\alpha(n,m)$ 使得 $P\big(F_{n,m} \geq F_\alpha(n,m)\big) = \alpha$.
- **定义 3.4** 我们称 z_α, $\chi^2_\alpha(n)$, $t_\alpha(n)$ 和 $F_\alpha(n,m)$ 分别为 $N(0,1)$, $\chi^2(n)$, $t(n)$ 和 $F(n,m)$ 分布的**上 α 分位数**, 并统一称为上 α 分位数.

- 容易理解, 上 α 分位数是 α 的减函数.
- 对于上 α 分位数, 容易验证 (参考后面的图)

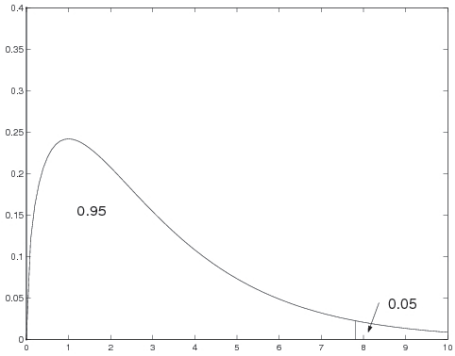
$$
\begin{aligned}
&P(Z \le z_\alpha) = 1-\alpha,\\
&P(\chi_n^2 \le \chi_\alpha^2(n)) = 1-\alpha,\\
&P(T_n \le t_\alpha(n)) = 1-\alpha,\\
&P(F_{n,m} \le F_\alpha(n,m)) = 1-\alpha.
\end{aligned}
$$

标准正态分布上分位数的图形



$\alpha = 0.05,\ z_\alpha = 1.645$

卡方分布上分位数的图形



$\alpha = 0.05,\ \chi_\alpha^2(3) = 7.815$

分位数的计算

- 对某些固定的 α, 可以查书后面的表得到 z_α, $\chi_\alpha(n)$, $t_\alpha(n)$ 和 $F_\alpha(n,m)$,
- 也可以用 Matlab、R、SAS 等直接计算.
- 记住下面的 (3.4) 和 (3.5) 式是有帮助的.

例 3.2

- 对 $Z \sim N(0,1)$, $T_n \sim t(n)$, 有 (参考后面的图)

$$P(\,|Z| \geq z_{\alpha/2}\,) = \alpha, \;\; P(\,|Z| \leq z_{\alpha/2}\,) = 1-\alpha. \tag{3.4}$$

$$P(\,|T_n| \geq t_{\alpha/2}(n)\,) = \alpha, \;\; P(\,|T_n| \leq t_{\alpha/2}(n)\,) = 1-\alpha. \tag{3.5}$$

- **证** 利用 $-Z \sim N(0,1)$, 得到

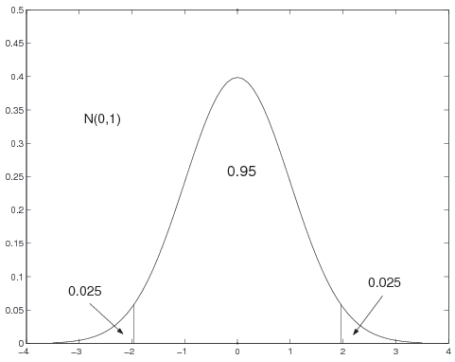
$$\begin{aligned} P(\,|Z| \geq z_{\alpha/2}\,) =& P(\,Z \geq z_{\alpha/2}) + P(\,-Z \geq z_{\alpha/2}) \\ =& \alpha/2 + \alpha/2 = \alpha. \end{aligned}$$

- $$P(\,|Z| \leq z_{\alpha/2}\,) = 1 - P(\,|Z| \geq z_{\alpha/2}\,) = 1-\alpha.$$

- (3.4) 的证明只用到标准正态密度关于原点的对称性.

- 由于 $t(n)$ 分布的密度函数也是关于原点对称的, 所以可以相同地证明 (3.5).

双侧分位数图形



$\alpha = 0.05$, $z_{0.05/2} = 1.96$

例 3.3(两个常用的正态分位数)

- $z_{0.025} = 1.96$, $z_{0.05} = 1.645$.

- **解** 查正态分布表可以得到以上结果.

- $z_{0.025} = 1.96$ 和 $z_{0.05} = 1.645$ 是两个最常用的分位数, 值得牢记.

例 3.4(F 分布的分位数)

- 对 $F(n,m)$ 的上 α 分位数 $F_\alpha(n,m)$, 有

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

- **证** 对 $F_{n,m}\sim F(n,m)$, 从定理 3.7 得到 $1/F_{n,m}\sim F(m,n)$.

- 于是对 $F_{m,n}\sim F(m,n)$, 得到

$$\begin{aligned}&P(F_{m,n}\geq 1/F_{1-\alpha}(n,m))\\=&P(1/F_{n,m}\geq 1/F_{1-\alpha}(n,m))\\=&P(F_{n,m}\leq F_{1-\alpha}(n,m))\\=&1-(1-\alpha)=\alpha.\end{aligned}$$

二项分布的分位数

- **定义 3.5** 设 $X\sim \mathrm{B}(n,p)$, 对于 $\alpha\in(0,1)$, 如果正整数 $B_\alpha(n,p)$ 使得

$$P\big(X\geq B_\alpha(n,p)\big)\leq\alpha,\ \ P\big(X\geq B_\alpha(n,p)-1\big)>\alpha, \tag{3.6}$$

  就称 $B_\alpha(n,p)$ 为二项分布 $B(n,p)$ 的上 α 分位数.

- 可以查表得到上 α 分位数 $B_\alpha(n,p)$，或用软件计算。

7.4 正态总体的区间估计

区间估计

- 在独立同分布场合, 样本均值 $\bar{X}_n$ 和样本方差 S^2 分别是总体均值 μ 和总体方差 σ^2 的无偏估计和相合估计, 说明样本均值和样本方差都是不错的估计量.

- 它告诉我们, 在 n 比较大的时候, 真值 μ 就在 $\bar{X}_n$ 的附近, 真值 σ^2 就在 S^2 的附近.

- 但是到底有多近呢? n 多大就够了呢?

- 区间估计可以回答这一问题.

7.4.1 已知 σ 时, μ 的置信区间

已知 σ 时, μ 的置信区间—例 1

- **例 4.1** 为了得到鲜牛奶的冰点, 对鲜牛奶的冰点进行了 21 次独立重复测量, 得到数据如下 (单位: 摄氏度).

$$\begin{array}{cccccc} -0.541 & -0.545 & -0.543 & -0.554 & -0.547 & -0.543 \\ -0.538 & -0.548 & -0.552 & -0.544 & -0.551 & -0.547 \\ -0.542 & -0.545 & -0.552 & -0.551 & -0.548 & -0.543 \\ -0.552 & -0.535 & -0.546 & & & \end{array}$$

  已知测量 (仪器) 的标准差是 $\sigma = 0.0048$, 测量没有**系统偏差** (也就是说测量值 X 的数学期望等于牛奶的冰点), 估计牛奶的冰点是多少.

- **解** 牛奶的冰点是常数, 测量值的随机性由测量误差造成.

- 通常认为测量值 X 服从正态分布 $N(\mu, \sigma^2)$, $\mu = \mathrm{E}X$ 是牛奶的冰点, $\sigma = 0.0048$ 是测量的标准差.

- 对 $n = 21$, 容易从数据计算出 $\bar{X}_n = -0.546$, 这是对 μ 的估计.

- 真正的 μ 到底距离 $\bar{X}_n = -0.546$ 有多远呢?

- 用 X_i 表示第 i 次测量值, 则 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的 21 个样本值, $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

- 取 $\alpha = 0.05$, 则 $(1-\alpha) = 0.95$, $z_{\alpha/2} = 1.96$.

- 于是得到

$$\begin{aligned} & P(|\bar{X}_n - \mu| \le 1.96\sigma/\sqrt{n}) \\ =& P\Big(\frac{|\bar{X}_n - \mu|}{\sigma/\sqrt{n}} \le 1.96\Big) = 0.95. \end{aligned} \tag{4.1}$$

- 也就是以 0.95 的概率保证 $|\bar{X}_n - \mu| \le 1.96\sigma/\sqrt{n}$,

- 现在 $\bar{X}_n = -0.546$, $\sigma = 0.0048, n = 21$, 所以我们以 0.95 的概率保证

$$\begin{aligned} \mu \in& [\bar{X}_n - 1.96\sigma/\sqrt{n},\ \bar{X}_n + 1.96\sigma/\sqrt{n}] \\ =& [-0.5481,\ -0.5440]. \end{aligned} \tag{4.2}$$

- 我们称 $[-0.5481,\ -0.5440]$ 是 μ 的置信度为 0.95 的置信区间. 置信区间的长度是 $-0.5440 + 0.5481 = 0.0041$.

- 容易得到以下的结论: 如果 $x_1, x_2, \cdots, x_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, σ 已知, 则 μ 的置信度为 0.95 的置信区间是

$$[\bar{X}_n - 1.96\sigma/\sqrt{n},\ \bar{X}_n + 1.96\sigma/\sqrt{n}]. \tag{4.3}$$

- 同样可以计算 μ 的置信度为 0.99 的置信区间是

$$[\bar{X}_n - 2.576\sigma/\sqrt{n},\ \bar{X}_n + 2.576\sigma/\sqrt{n}], \tag{4.5}$$
$$2.576 = z_{0.01/2}.$$

置信区间定义

- **定义 4.1** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本,$\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$, θ 是未知参数, $\hat{\theta}_1 = \hat{\theta}_1(\boldsymbol{X})$, $\hat{\theta}_2 = \hat{\theta}_2(\boldsymbol{X})$ 是两个统计量. 对于给定的 $\alpha \in (0,1)$, 如果有

$$P(\,\hat{\theta}_1 \le \theta \le \hat{\theta}_2\,) \ge 1 - \alpha, \tag{4.6}$$

  就称 $[\hat{\theta}_1, \hat{\theta}_2]$ 为参数 θ 的 **置信度**为 $(1-\alpha)$ 的**置信区间** (confidence interval, CI).

- 在定义 4.1 中, 置信度又称为**置信水平** (confidence level), 置信区间的右端点 $\hat{\theta}_2$ 又称为**置信上界**, 置信区间的左端点 $\hat{\theta}_1$ 又称为**置信下界**.

- 由于 $\hat{\theta}_1 = \hat{\theta}_1(\boldsymbol{X})$ 和 $\hat{\theta}_2 = \hat{\theta}_2(\boldsymbol{X})$ 都是随机变量的函数, 因而是随机变量.

- 但是给定样本观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$, 就得到了一个具体的闭区间 $[\hat{\theta}_1(\boldsymbol{x}), \hat{\theta}_2(\boldsymbol{x})]$, 这个闭区间或者包含 μ 或者不包含 μ，我们对 $1-\alpha$ 置信度的理解是：

- 如果允许我们反复独立抽取样本得到多次置信区间，μ 落入这些置信区间中的比例接近于 $1-\alpha$。

- 很明显, 在相同的置信度下, 置信区间的长度越小越好.

正态总体 σ 已知时 μ 的置信区间

- **例 4.2** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, 当标准差 σ 已知, 均值 μ 的置信度为 $(1-\alpha)$ 的置信区间是

$$\Big[\,\bar{X}_n - \frac{z_{\alpha/2}\sigma}{\sqrt{n}},\ \bar{X}_n + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\,\Big]. \tag{4.7}$$

- **解** 这时
$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$
- 对于给定的置信度 $(1-\alpha)$, 有
$$\begin{aligned} &P\big(\bar{X}_n - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{X}_n + z_{\alpha/2}\sigma/\sqrt{n}\big) \\ =&P(|\bar{X}_n - \mu| \le z_{\alpha/2}\sigma/\sqrt{n}) \\ =&P(|Z| \le z_{\alpha/2}) = 1-\alpha. \end{aligned}$$
- 所以, 已知标准差 σ 时, 均值 μ 的置信度为 $(1-\alpha)$ 的置信区间是 (4.7).

正态总体 σ 已知时 μ 的置信区间—讨论

- 置信区间 (4.7) 的长度是 $L = 2z_{\alpha/2}\sigma/\sqrt{n}$.
- L 越小, 置信区间提供的信息越准确.
- 由于 $z_{\alpha/2}$ 是置信度 $(1-\alpha)$ 的增函数, 所以对于相同的标准差 σ, 从 (4.7) 看出以下结论:

  (1) 置信区间的中心总是样本均值;

  (2) 置信度 $1-\alpha$ 越高, 置信区间就越长;

  (3) 样本量 n 越大, 置信区间越短.
- 标准差 σ 越大，说明总体越分散，从中估计 μ 越难估计准确，置信区间越长。
- 为了克服置信区间过长带来的不足, 同时考虑置信度不能太低, 人们一般使用置信度为 $1-\alpha = 0.95$ 的置信区间.
- $z_{\alpha/2} = 1.96$ 是值得牢记的.

7.4.2 未知 σ 时 μ 的置信区间

未知 σ 时 μ 的置信区间

- 在例 4.2 中, 如果 σ 是未知数, 自然想到用样本标准差 S 代替 σ.
- 根据定理 3.6, 统计量
$$T_{n-1} = \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1).$$

- 对于给定的置信度 $(1-\alpha)$, 查 t 分布表 (附录 C2) 可以得到上分位数 $t_{\alpha/2}(n-1)$, 这时

$$P\Big(\frac{|\bar{X}_n-\mu|}{S/\sqrt{n}}\le t_{\alpha/2}(n-1)\Big)=P(|T_{n-1}|\le t_{\alpha/2}(n-1))=1-\alpha.$$

- 由于

$$\begin{aligned}&\Big\{\frac{|\bar{X}_n-\mu|}{S/\sqrt{n}}\le t_{\alpha/2}(n-1)\Big\}\\=&\Big\{\bar{X}_n-\frac{t_{\alpha/2}(n-1)S}{\sqrt{n}}\le\mu\le\bar{X}_n+\frac{t_{\alpha/2}(n-1)S}{\sqrt{n}}\Big\},\end{aligned}$$

- 所以在置信度 $(1-\alpha)$ 下, μ 的置信区间是

$$\Big[\bar{X}_n-\frac{t_{\alpha/2}(n-1)S}{\sqrt{n}},\ \bar{X}_n+\frac{t_{\alpha/2}(n-1)S}{\sqrt{n}}\Big]. \tag{4.8}$$

例 4.3

- 在例 4.1 中, 假设标准差 σ 未知, 计算均值 μ 的置信度为 0.95 的置信区间.
- **解** 从例 4.1 中的数据可以计算出样本标准差 $S=0.005$, 查表得到 $t_{\alpha/2}(20)=2.086$.
- 代入 (4.8), 得到 μ 的置信区间

$$\begin{aligned}&[-0.546-2.086\times0.005/\sqrt{21},\ -0.546+2.086\times0.005/\sqrt{21}]\\=&[-0.5483,-0.5437].\end{aligned}$$

- 置信区间的长度是 0.0046.
- 这个置信区间和已知 $\sigma=0.0048$ 的置信区间基本相同. 这是因为样本量 $n=21$ 时, $S\approx\sigma$, $t(20)$ 的密度和 $N(0,1)$ 的密度也基本相同的原因.

例 4.4

- 在例 4.1 中, 只使用前 7 个数据计算 μ 的置信度为 0.95 的置信区间和置信区间的长度.

  (1) 已知标准差 $\sigma=0.0048$;

  (2) 未知标准差 σ.

- **解** $(1-\alpha)=0.95, \alpha=0.05$. 前 7 个数据的样本均值 $\bar{X}_7=-0.5444$, 样本标准差 $S=0.0051$.

- (1) 已知 $\sigma=0.0048$ 时, 利用 $z_{\alpha/2}=1.96$ 和公式 (4.7) 得到 μ 的置信区间

$$\begin{aligned}&[-0.5444-1.96\times 0.0048/\sqrt{7},\\&-0.5444+1.96\times 0.0048/\sqrt{7}]\\=&[-0.5480,\ -0.5408].\end{aligned}$$

  置信区间的长度是 0.0072.

- (2) 未知 σ 时, 查 t 分布表, 得到 $t_{\alpha/2}(6)=2.447$, 代入公式 (4.8) 得到 μ 的置信度为 0.95 的置信区间

$$\begin{aligned}&[-0.5444-2.447\times 0.0051/\sqrt{7},\\&-0.5444+2.447\times 0.0051/\sqrt{7}]\\=&[-0.5491,\ -0.5397].\end{aligned}$$

  置信区间的长度是 0.0094.

- 例 4.1 中置信区间长度为 0.0041.

- 只使用 7 个数据时的置信区间比使用 21 个数据的置信区间要长一些, 说明样本量越大, 置信区间越精确.

7.4.3 方差 σ^2 的置信区间

方差 σ^2 的置信区间

- **例 4.5** 地球生物的演变经历了漫长的岁月, 只有化石为这一演变进行了记录. 现在科学家们利用物质的放射性衰变来研究生物的演变规律.

- 几乎所有的矿物质含有 K(钾) 元素及其同位素 ^{40}K(钾 40). ^{40}K 并不稳定, 它可以缓慢地衰变成 ^{40}Ar(氩 40) 和 ^{40}Ca(钙 40).

- 于是知道了 ^{40}K 的衰变速率, 就可以通过测量化石中的 ^{40}K 和 ^{40}Ar 的比例 (钾氩比) 估计化石的形成年代.

- 下面是根据钾氩比计算出的德国黑森林中发掘的 19 个化石样品的形成年龄 (单位: 百万年). (见 [4])

$$249\ \ 254\ \ 243\ \ 268\ \ 253\ \ 269\ \ 287\ \ 241\ \ 273$$
$$306\ \ 303\ \ 280\ \ 260\ \ 256\ \ 278\ \ 344\ \ 304\ \ 283\ \ 310.$$

- 假设每个样品的估算年代都服从正态分布. 为了评价钾氩比方法的估计精度, 需要完成以下工作.

  (1) 计算 σ^2 的置信度为 $(1-\alpha)$ 的置信区间;

  (2) 计算 σ^2 的置信度为 0.95 的置信区间;

  (3) 计算标准差 σ 的置信度为 0.95 的置信区间.

- **解** (1) 用 X_i 表示第 i 个样品的估算年代, 则根据题意, $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, 样本量 $n = 19$.

- 样本方差

$$S^2 = \frac{1}{(n-1)} \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$$

  是总体方差 σ^2 的无偏估计和强相合估计.

- 于是

$$\frac{(n-1)S^2}{\sigma^2}$$

  应当在 $(n-1)$ 附近.

- 又从定理 3.4 知道,

$$\chi^2_{n-1} \triangleq \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

  记 $\lambda_1 = \chi^2_{1-\alpha/2}(n-1)$, $\lambda_2 = \chi^2_{\alpha/2}(n-1)$, 则

$$P\Big(\lambda_1 \le \frac{(n-1)S^2}{\sigma^2} \le \lambda_2\Big) = 1 - \alpha.$$

- 对不等式进行变换得

$$\Big\{\lambda_1 \le \frac{(n-1)S^2}{\sigma^2} \le \lambda_2\Big\} = \Big\{\frac{(n-1)S^2}{\lambda_2} \le \sigma^2 \le \frac{(n-1)S^2}{\lambda_1}\Big\}$$

- 所以在置信度 $(1-\alpha)$ 下, σ^2 的置信区间是

$$\Big[\frac{(n-1)S^2}{\lambda_2},\ \ \frac{(n-1)S^2}{\lambda_1}\Big]. \tag{4.9}$$

- (2) 本例中 $n-1=18$, $\alpha/2=0.025$.
- 可以计算出
$$\bar{X}_n = 276.9, \;\; S^2 = 733.4.$$
- 查 $\chi^2(18)$ 表得到
$$\lambda_2 = \chi^2_{0.025}(18) = 31.53, \;\; \lambda_1 = \chi^2_{1-0.025}(18) = 8.23.$$
- 将上述数据代入 (4.9), 得到 σ^2 的置信度为 0.95 的置信区间
$$\begin{aligned}&\Big[\frac{18\times 733.4}{31.53}, \; \frac{18\times 733.4}{8.23}\Big]\\ =&[418.7, \; 1604.0] \text{ (百万年)}^2.\end{aligned}$$
- (3) 由于
$$\begin{aligned}&\Big\{\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}} \le \sigma \le \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\Big\}\\ =&\Big\{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\Big\}. \qquad (4.10)\end{aligned}$$
- 所以 σ 的置信度为 0.95 的置信区间是
$$\Big[\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}, \; \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\Big]. \qquad (4.11)$$
- 代入数值后得到 σ 的置信度为 0.95 的置信区间是
$$[\sqrt{418.7}, \sqrt{1604}] = [20.5, \; 40.05] \;\; \text{(百万年)}.$$
- **注:** 从图 4.1 看出, 在置信水平 0.95 下, 置信区间 (4.9) 的长度不是最短的一个, 但是它有计算和使用方便的优点.

7.4.4 均值差 $\mu_1-\mu_2$ 的置信区间

均值差 $\mu_1-\mu_2$ 的置信区间

- 设 $X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$. $X_1,X_2,\cdots,X_n$ 是来自 X 的样本, $Y_1,Y_2,\cdots,Y_m$ 是来自 Y 的样本, 设总体 X 和总体 Y 独立, 于是
$$X_1,X_2,\cdots,X_n,Y_1,Y_2,\cdots,Y_m$$
相互独立.

- 下面构造 $\mu_1-\mu_2$ 置信区间.
- 用 $\bar{X}_n$, $\bar{Y}_m$ 分别表示样本均值, 用 S_1^2, S_2^2 分别表示样本方差,
- 则
$$\bar{X}_n-\bar{Y}_m\sim N(\mu_1-\mu_2,\sigma_1^2/n+\sigma_2^2/m).$$
- 于是得到
$$Z=\frac{(\bar{X}_n-\bar{Y}_m)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}\sim N(0,1). \tag{4.12}$$

- (1) 已知 σ_1^2, σ_2^2 时, 对置信度 $(1-\alpha)$, 利用 (4.12) 构造出 $\mu_1-\mu_2$ 的置信区间
$$\Big[\ (\bar{X}_n-\bar{Y}_m)-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}},$$
$$(\bar{X}_n-\bar{Y}_m)+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\ \Big]. \tag{4.13}$$
- (2) 未知 σ_1^2,σ_2^2, 但是已知 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时, 由
$$\xi_1=\frac{(n-1)S_1^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{j=1}^{n}(X_j-\bar{X}_n)^2\sim\chi^2(n-1);$$
$$\xi_2=\frac{(m-1)S_2^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{j=1}^{m}(Y_j-\bar{Y}_m)^2\sim\chi^2(m-1), \tag{4.14}$$
及 ξ_1, ξ_2 独立
- 得到
$$\xi_1+\xi_2\sim\chi^2(n+m-2).$$
- 引入
$$S_W^2=\frac{(\xi_1+\xi_2)\sigma^2}{n+m-2}=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}, \tag{4.15}$$
- 根据定理 3.4 知道 Z, ξ_1, ξ_2 独立, 再注意 $\sigma_1^2=\sigma_2^2=\sigma^2$ 和利用定理 3.5 得到
$$\frac{(\bar{X}_n-\bar{Y}_m)-(\mu_1-\mu_2)}{S_W\sqrt{1/n+1/m}}$$
$$=\frac{Z}{\sqrt{(\xi_1+\xi_2)/(n+m-2)}}\ \sim\ t(n+m-2). \tag{4.16}$$

- 利用 (4.16) 构造出 $\mu_1-\mu_2$ 的置信度为 $(1-\alpha)$ 的置信区间

$$\Big[\ (\bar{X}_n-\bar{Y}_m)-t_{\alpha/2}S_W\sqrt{\frac{1}{n}+\frac{1}{m}},\\ (\bar{X}_n-\bar{Y}_m)+t_{\alpha/2}S_W\sqrt{\frac{1}{n}+\frac{1}{m}}\Big]. \tag{4.17}$$

其中 $t_{\alpha/2}=t_{\alpha/2}(n+m-2)$.

7.4.5 方差比 σ_1^2/σ_2^2 的置信区间

方差比 σ_1^2/σ_2^2 的置信区间

- 在 D 的假设下, 我们计算 σ_1^2/σ_2^2 的置信区间.
- 根据定理 3.4 和定理 3.7, 知道

$$F=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n-1,m-1).$$

- 利用 F_α 表示 $F_\alpha(n-1,m-1)$, 得到

$$P\Big(F_{1-\alpha/2}\le\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\le F_{\alpha/2}\Big)=1-\alpha.$$

- 于是在置信度 $(1-\alpha)$ 下, 可以得到 σ_1^2/σ_2^2 的置信区间

$$\Big[\ \frac{S_1^2}{S_2^2F_{\alpha/2}},\ \frac{S_1^2}{S_2^2F_{1-\alpha/2}}\ \Big].$$

7.4.6 单侧置信区间

单侧置信区间

- **定义 4.2** 设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$, θ 是未知参数, $\overline{\theta}=\overline{\theta}(\boldsymbol{X})$, $\underline{\theta}=\underline{\theta}(\boldsymbol{X})$ 是两个统计量. 对于给定的 $\alpha\in(0,1)$,

  1. 如果有

  $$P(\,\theta\le\overline{\theta}\,)\ge 1-\alpha, \tag{4.18}$$

  就称 $\overline{\theta}$ 为参数 θ 的置信度为 $(1-\alpha)$ 的**单侧置信上限**;

  2. 如果有

  $$P(\,\theta\ge\underline{\theta}\,)\ge 1-\alpha, \tag{4.19}$$

  就称 $\underline{\theta}$ 为参数 θ 的置信度为 $(1-\alpha)$ 的**单侧置信下限**.

正态总体 μ 和 σ^2 的单侧置信限表

条件	μ的单侧置信上限	μ 的单侧置信下限
已知 σ	$\overline{\mu} = \bar{X}_n + z_\alpha \sigma/\sqrt{n}$	$\underline{\mu} = \bar{X}_n - z_\alpha \sigma/\sqrt{n}$
未知 σ	$\overline{\mu} = \bar{X}_n + t_\alpha(n-1)S/\sqrt{n}$	$\underline{\mu} = \bar{X}_n - t_\alpha(n-1)S/\sqrt{n}$
条件	σ^2 的单侧置信上限	σ^2 的单侧置信下限
未知 μ	$\overline{\sigma}^2 = (n-1)S^2/\chi^2_{1-\alpha}(n-1)$	$\underline{\sigma}^2 = (n-1)S^2/\chi^2_\alpha(n-1)$

(4.20)

其他的单侧置信限可以类似得到, 见正态总体和正态逼近置信区间表.

7.5 非正态总体和比例 p 的置信区间

7.5.1 正态逼近法

正态逼近法—σ 已知

- 总体分布不是正态分布的情况也需要对均值和方差计算置信区间.
- 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, $\mu = \mathrm{E}X$, $\sigma^2 = \mathrm{Var}(X)$ 分别是总体均值和总体方差.
- 根据中心极限定理, 对较大的样本量 n,

  $$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

  近似服从标准正态分布.
- 对较大的 n, 有

  $$P\big(|\bar{X}_n - \mu| \le z_{\alpha/2}\sigma/\sqrt{n}\big) \approx (1-\alpha).$$

- 于是, 已知标准差 σ 时, 对置信度 $(1-\alpha)$, 总体均值 μ 的近似置信区间仍然是

  $$\Big[\bar{X}_n - \frac{z_{\alpha/2}\sigma}{\sqrt{n}},\ \bar{X}_n + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\Big].$$

正态逼近法—σ 未知

- 当 σ 未知时, 对较大的 n, S 是 σ 的强相合估计, 所以可以用 S 代替 σ.
- 根据例 3.1
$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim N(0,1)$$
近似成立.
- 所以未知标准差 σ 时, 均值 μ 的置信度为 $(1-\alpha)$ 的近似置信区间是
$$\Big[\bar{X}_n - \frac{z_{\alpha/2}S}{\sqrt{n}},\ \bar{X}_n + \frac{z_{\alpha/2}S}{\sqrt{n}}\Big]. \tag{5.1}$$
- n 越大, 近似的程度越好. 我们称以上方法为**正态逼近法**.
- 使用正态逼近法时，一般的要求是 $n \geq 30$.

例 5.1

- 研究年龄和血液中的各种成份之间的关系.
- 通过随机抽样调查了 30 个 30 岁健康公民的血小板数.
- 数据如下 (单位: 万/mm^3):

26 19 18 16 26 17 20 20 19 22 19 12 29 15 22
19 27 25 28 24 35 28 19 23 31 30 23 30 17 22

- 用 μ 表示 30 岁健康公民的血小板数的总体均值. 对于置信度 $(1-\alpha) = 0.95$, 计算 μ 的置信区间.
- **解** 可以认为被选到的个体的血小板数是独立同分布的. 经过计算得到 $\bar{X}_n = 22.7$, $S = 5.45$. 代入 (5.1) 得到置信度 0.95 下, μ 的近似置信区间
$$\Big[27.7 - 1.96 \times 5.45/\sqrt{30},\ 22.7 + 1.96 \times 5.45/\sqrt{30}\Big]$$
$$= [25.75,\ 29.65].$$

7.5.2 比例 p 的置信区间

比例 p 的置信区间

- **例 5.2** 设 $X_1, X_2, \cdots, X_n$ 是来自两点分布总体 $B(1,p)$ 的样本, 对置信度 $(1-\alpha)$, 当 n 较大 (至少使得 $5 < n\hat{p} < n-5$), p 的近似置信区间是

$$\Big[\frac{b-\sqrt{b^2-4ac}}{2a},\ \ \frac{b+\sqrt{b^2-4ac}}{2a}\Big], \tag{5.2}$$

  其中

$$a = 1+\frac{z_{\alpha/2}^2}{n},\ \ b = 2\bar{X}_n+\frac{z_{\alpha/2}^2}{n},\ \ c = \bar{X}_n^2.$$

- 证明略。

- 当 n 更大, 更简单一些的近似置信区间是

$$\Big[\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}},\ \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\Big]. \tag{5.3}$$

- (5.3) 式是由于

$$\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)/n}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$$

估计 p 所需的样本量

- **例 5.3** 给定置信度 $(1-\alpha)$, 要使得置信区间 (5.2) 或 (5.3) 的长度不超过 d, 只要取样本量

$$n \geq \Big(\frac{z_{\alpha/2}}{d}\Big)^2. \tag{5.7}$$

- 取 $(1-\alpha) = 0.95$ 时, 可以列出和 d 对应的 n 如下.

$d=$	0.14	0.12	0.10	0.08	0.06	0.04	0.02	0.01
$n=$	196	267	385	601	1068	2401	9604	38416

(5.8)

- 证明略。

例 5.4

- 饮用水资源的匮乏限制了我国许多城市的经济发展. 为了节约用水, 城市甲准备对自来水提价. 现在需要对每吨水提价 0.5 元还是 0.8 元进行随机抽样调查, 为的是即达到节水的目的, 又不影响百姓的日常生活.

(1) 用 p 表示赞同提价 0.5 元的人口比例, 为了得到 p 的置信度为 $(1-\alpha)=0.95$ 的置信区间, 且置信区间长度不超过 0.04, 应当随机抽样调查多少人?

(2) 如果随机抽样调查的 $n=2500$ 个人中有 1668 个人同意提价 0.5 元, 计算 p 的置信度为 0.95 的置信区间,

(3) 计算 (2) 中置信区间的长度.

- **解** (1) $d=0.04$, $\alpha=0.05$. 从 (5.8) 知道至少应当调查 2401 个人.
- (2) 2500 个人中有 1668 个人同意提价 0.5 元时, 可以计算出

$$\bar{X}_n=\frac{1668}{2500}=0.6672,\ n=2500,\ z_{\alpha/2}=1.96.$$

由于样本量已经较大, $\bar{X}_n$ 偏离 0.5 又不远, 我们使用简单一些的置信区间 (5.3).

- 将上面的数代入 (5.3), 得到 p 的置信度为 0.95 的置信区间

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

$$=[0.6487,0.6857].$$

于是, 我们以 95/% 的把握保证, 赞同提价 0.5 元的人口比例在 64/% 至 69/% 之间.

- (3) 置信区间的长度是 0.037.

第八章　假设检验

8.1　假设检验的概念

假设检验引入

- 假设检验是统计推断的一个主要部分.
- 其想法和前面的最大似然类似：如果实际观测到得到数据在某假设下不太可能出现则认为该假设错误。
- **例 1.1**　一条新建的南北交通干线全长 10 公里. 公路穿过一个隧道 (长度忽略不计), 隧道南面 3.5 公里, 北面 6.5 公里. 在刚刚通车的一个月中, 隧道南发生了 3 起交通事故, 而隧道北没有发生交通事故, 能否认为隧道南的路面更容易发生交通事故?
- **解**　隧道将公路分为两段, 隧道南 3.5 公里, 隧道北 6.5 公里.
- 用 p 表示一起交通事故发生在隧道南的概率, 则 $p = 0.35$ 表示隧道南北的路面发生交通事故的可能性相同. $p > 0.35$ 表示后隧道南的路面发生交通事故的概率比隧道北的路面发生交通事故的概率大.
- 为了作出正确的判断, 先作一个假设

  $$H_0 : \ p = 0.35.$$

  我们称 H_0 是**原假设** 或 **零假设**.
- 再作一个**备择假设**

  $$H_1 : \ p > 0.35.$$

- 在本问题中, 如果判定 H_0 不对, 就应当承认 H_1.
- 三起交通事故的发生是相互独立的, 他们之间没有联系.

- 如果 H_0 为真, 则每一起事故发生在隧道南的概率都是 0.35, 于是这三起交通事故都发生在隧道南的概率是

$$P = 0.35^3 \approx 0.043.$$

- 这是一个很小的概率, 一般不容易发生.

- 所以我们否定 H_0, 认为隧道南的路面发生交通事故的概率比隧道北大.

- 做出以上结论也有可能犯错误, 犯错误的概率正是 0.043.

- 这是因为当隧道南北的路面发生交通事故的概率相同, 而 3 起交通事故又都出现在隧道南时, 我们才犯错误. 这一概率正是 $P = 0.043$.

- 于是, 我们判断正确的概率是 $1 - 0.043 = 95.7\%$(在多次解决类似问题意义下).

- 通过对例 1.1 的分析, 我们得到以下的概念.

- 进行假设检验时, 根据问题的背景, 先作出原假设

$$H_0 : \ p = 0.35,$$

  及其备择假设

$$H_1 : \ p > 0.35.$$

- 然后在 H_0 的下, 计算出观测数据出现的概率 P.

- 如果 P 很小 (一般用 0.05 衡量), 就应当否定 H_0, 承认 H_1;

- 如果 P 不是很小, 也不必急于承认 H_0, 这是因为证据往往还不够充分. 如果继续得到的观测数据还不能使得 P 降低下来, 再承认 H_0 不迟.

- 为了简便, 我们把以上的原假设和备择假设记作

$$H_0 : \ p = 0.35 \quad vs \quad H_1 : \ p > 0.35.$$

  其中的 vs 是 versus 的缩写.

假设检验的一般提法

- 一般来讲, 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, θ 是总体 X 的未知参数, 但是已知 $\theta \in \Theta_0 + \Theta_1$.
- 这里 Θ 是 θ 的大写, Θ_0, Θ_1 是互不相交的参数集合.
- 对于假设

  $$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

  的检验法 W(我们用 W 表示这个检验法), 如果否定 H_0 时犯错误的概率不超过 α, 就称 W 是检验水平为 α 的检验, 称 α 是检验法 W 的检验水平.
- 检验法 W 可以被事件 W 完全确定，事件 W 发生时拒绝 H_0，称 W 为**拒绝域**。
- **定义 1.1** 设 α 是 $(0,1)$ 中的常数. 如果对一切的 $\theta \in \Theta_0$, 有

  $$P_\theta(W) \leq \alpha,$$

  就称拒绝域 W 的检验水平或 显著性水平是 α.

假设检验的两类错误

- 在解决假设检验的问题时, 无论作出否定还是接受原假设 H_0 的决定, 都有可能犯错误.
- 我们称否定 H_0 时犯的错误为第一类错误, 接受 H_0 时犯的错误为第二类错误.
		检验结果	
		H_0	H_1
真实情况	H_0	正确	第 I 类错误
	H_1	第 II 类错误	正确
- 假设检验一般控制第一类错误在检验水平 α 以下，所以否定 H_0 时结论比较可靠。
- 如果承认 H_0，可能犯第二类错误，错误概率可能会比较大。
- 在正确的统计推断前提下, 犯错误的原因总是随机因素造成的.
- 要有效减少犯错误的概率, 只好增加观测数据，或在可能的情况下提高数据的质量, 这相当于降低数据的样本方差.

- 在例 1.1 中, 如果第一起交通事故发生后, 就断定隧道南更容易发生交通事故, 犯第一类错误的概率是 0.35.
- 当第二起交通事故发生后, 断定隧道南更容易发生交通事故, 犯第一类错误的概率是 $0.35^2 = 0.1225$.
- 当第三起交通事故发生后, 断定隧道南更容易发生交通事故, 犯第一类错误的概率是 $0.35^3 = 0.043$.
- 如果第四起交通事故又发生在隧道南, 否定 $p = 0.35$ 时犯第一类错误的概率是 $0.35^4 = 0.015$.

例 1.2: 第一类错误与第二类错误的比较

- 一个有 20 多年教龄的教师声称他上课从来不“点名”. 如何判定他讲的话是真实的?
- 为了解决这个问题, 我们也做一个原假设 H_0: 他没有点过名, 然后再调查 H_0 是否为真.
- 当调查了他教过的 3 个班, 都说他没有点过名, 这时如果承认 H_0, 犯错误的概率还是较大的.
- 当调查了他教过的 10 个班, 都说他没有点过名, 这时承认 H_0 犯错误的概率会明显减少.
- 如果调查了他教过的 30 个班, 都说他没有点过名, 这时承认 H_0 犯错误的概率就会很小了.
- 可惜调查 30 个班是很难做到的.
- 反过来, 在调查中只要有人证实这位老师点过名, 就可以否定 H_0 了 (不论调查了几个班), 并且这样做犯错误的概率很小.
- 例 1.2 告诉我们, 要否定原假设 H_0 是比较简单的, 只要观测到了 H_0 下小概率事件就可以。
- 要承认 H_0 就比较费力了: 必须有足够多的证据 (样本量), 才能够以较大的概率保证 H_0 的真实.
- 在这个例子中还有一个现象值得注意: 当调查 10 个班发现都没有点过名就承认 H_0 时, 即使判断失误, 造成的后果也不严重. 因为数据已经说明这位老师不爱点名.

8.2 正态均值的假设检验

8.2.1 已知 σ 时, μ 的正态检验法

已知 σ 时, μ 的正态检验法—例 2.1

- **例 2.1** 一台方差是 0.8克^2 的自动包装机在流水线上包装净重 500 克的袋装白糖. 现在随机抽取了该包装机包装的 9 袋白糖, 测得净重如下 (单位: 克).

$$\begin{array}{ccccc} 499.12 & 499.48 & 499.25 & 499.53 & 500.82 \\ 499.11 & 498.52 & 500.01 & 498.87. & \end{array}$$

  能否认为包装机在正常工作?

- 分析: 抽查的 9 袋白糖中有 7 袋净重少于 500 克, 似乎已经说明 $\mu_0 = 500$ 不对.

- 但是, 由于包装机的方差是 0.8 克, 所以也有可能是由于包装机的随机误差导致了以上的数据.

- 下面的分析说明, 由于随机误差导致上述观测数据的概率不超过 0.05.

- **解** 我们将包装机包装的袋装白糖的净重视为总体 X, 则 $X \sim N(\mu, \sigma^2)$, 其中 $\sigma^2 = 0.8$ 已知，μ 未知.

- 用 X_j 表示第 j 袋白糖的净重, 则 $X_1, X_2, \cdots, X_9$ 是来自总体 X 的 $n = 9$ 个样本.

- 设 $\mu_0 = 500$, 作假设

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0.$$

- 在 H_0 下,

$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{9}} \sim N(0, 1).$$

- $|Z|$ 取值应当与 0 差距不大. 当 $|Z|$ 取值较大时, 要否定 H_0.

- 对于标准正态分布的上 $\alpha/2$ 分位数 $z_{\alpha/2}$, 在 H_0 下,

$$P(|Z| \geq z_{\alpha/2}) = \alpha,$$

- 如果取 $\alpha = 0.05$, 则 $z_{\alpha/2} = 1.96$, $P(|Z| \geq 1.96) = 0.05$.

- 当 $|Z| \geq 1.96$ 时, 不该发生的小概率事件发生了, 于是否定原假设 H_0.
- 本例中 $\bar{X}_n = 499.412$,

$$|Z| = \Big|\frac{499.412 - 500}{\sqrt{0.8/9}}\Big| = 1.97 > 1.96,$$

- 所以应当否定 H_0, 认为包装机没有正常工作.

- 在例 2.1 中, 称 α 为**检验的显著性水平**, 简称为显著性水平, 检验水平, 或水平 (level);
- 称 Z 为**检验统计量**;
- 称 $\{|Z| \geq z_{\alpha/2}\}$ 为检验的**拒绝域**或否定域;
- 如果 $\{|Z| \geq z_{\alpha/2}\}$ 发生, 就称**检验是显著的**.
- 这时否定 H_0, 犯第一类错误的概率不超过 α; 检验水平就是犯第一类错误的概率.
- 值得注意, 拒绝域 $\{|Z| \geq z_{\alpha/2}\}$ 是一个事件, 它的发生与否由 $|Z|$, 从而由观测样本 $X_1, X_2, \cdots, X_n$ 决定.

已知 σ 时, μ 的正态检验法

- 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, σ 已知时,

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

的显著性水平为 α 的拒绝域是

$$W = \{|Z| \geq z_{\alpha/2}\}, \text{ 其中 } Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}. \tag{2.1}$$

- 如果 $|Z| \geq z_{\alpha/2}$ 发生, 就称检验是显著的, 表示结论和假设有显著性差异.
- 这时, 否定 H_0 犯错误的概率不超过 α.
- 特别当 $\alpha = 0.05$, $z_{\alpha/2} = 1.96$.
- 由于这种检验方法是基于正态分布的方法, 所以又称为**正态检验法**或 Z 检验法.

- 在例 2.1 中, 如果取检验水平 $\alpha = 0.04$, 则临界值 $z_{\alpha/2} = z_{0.02} = 2.054$ (查附录 C1(续)).
- 这时 $|Z| = 1.97 < 2.054$, 不能否定 H_0.
- 说明在不同的检验水平下可以得到不同的检验结果.
- 降低犯第一类错误的概率, 就会使得拒绝域减小:

$$\Big\{|Z| > z_{0.04/2}\Big\} \subset \Big\{|Z| > z_{0.05/2}\Big\}.$$

从而拒绝 H_0 的机会变小，接受 H_0 的机会变大。

8.2.2 p 值检验法

p 值检验法

- 在例 2.1 中, 已知样本标准差 $\sigma = 0.8$, 从实际数据计算得到 $|z| = 1.97$.
- 如果把拒绝域取成

$$W = \{|Z| \geq 1.97\}$$

则刚刚能够拒绝 H_0.

- 这时犯第一类错误的概率是

$$p = P(|Z| \geq 1.97) = 2[1 - \Phi(1.97)] = 0.0488.$$

- 我们称 $p = 0.0488$ 是**检验的 p 值** ($p-$value).
- P 值越小, 数据提供的否定 H_0 的证据越充分.
- p 值是在 H_0 成立的假设下观测到的样本倾向于 H_1 的概率。
- 如果检验的显著性水平 α 是事先给定的, 当 P 值小于等于 α, 就要否定 H_0.
- 检验法 (2.1) 的 P 值是

$$P = P(\,|Z| \geq |z|\,) = 2\Phi(-|z|), \text{ 其中 } z = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}.$$

8.2.3 未知 σ 时, 均值 μ 的 t 检验法

未知 σ 时, 均值 μ 的 t 检验法—例 2.2

- **例 2.2** 在例 2.1 中如果 9 个袋装白糖的样品是从超级市场仓库中随机抽样得到的, 能否认为这批 500 克袋装白糖的平均重量是 500 克?
- **解** σ 未知, 需要寻找其他的方法.
- 对 $\mu_0 = 500$ 克, 仍作假设

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0.$$

- 在标准差 σ 未知时, 可用样本标准差

$$S = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2}$$

代替 σ.

- 在 H_0 下, 从 §7.3 的定理 3.6 知道检验统计量

$$T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t(n-1).$$

说明 T 在 0 附近取值是正常的, 如果 $|T|$ 取值较大就应当拒绝 H_0.

- 根据分位数 $t_{\alpha/2}(n-1)$ 的性质, 有

$$P(|T| \geq t_{\alpha/2}(n-1)) = \alpha.$$

- 于是 H_0 的显著性水平为 α 的拒绝域是

$$\left\{ |T| \geq t_{\alpha/2}(n-1) \right\}, \ T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}. \tag{2.2}$$

- 现在 $\mu_0 = 500$. 取 $\alpha = 0.05$, 查表得到 $t_{0.05/2}(8) = t_{0.025}(8) = 2.306$. 经过计算得到 $\bar{X}_n = 499.412$, $S = 0.676$,

$$|T| = \left| \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \right| = 2.609 > 2.306.$$

- $|T|$ 大于临界值 2.306, 所以应当否定 H_0, 认为 $\mu_0 \neq 500$.
- 在本例中, $\bar{X}_n = 499.412 < 500$, 所以认为供应的白糖是缺斤少两的. 作出以上判断也有可能犯错误, 但是犯错误的概率不超过 $\alpha = 0.05$.

- 我们将例 2.2 中的方法总结如下: 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, σ 未知时,

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

的显著性水平为 α 的拒绝域是

$$\big\{|T| \geq t_{\alpha/2}(n-1)\big\}, \text{ 其中 } T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}. \tag{2.3}$$

- 如果 $|T| \geq t_{\alpha/2}(n-1)$ 发生, 就称检验是显著的, 这时否定 H_0 犯错误的概率不超过 α.
- 由于这种检验方法是基于 t 分布的方法, 所以又称为t **检验法**.
- 设 T 统计量的计算结果为 a，则检验法 (2.3) 的 P 值为

$$\begin{aligned} P =& P\Big(|T_{n-1}| \geq |a| \Big) \\ =& 2P\Big(T_{n-1} \geq |a| \Big), \text{ 其中 } T_{n-1} \sim t(n-1). \end{aligned}$$

其中的 $\bar{x}_n$, s 分别是实际计算出的样本均值, 样本标准差.

8.2.4 未知 σ 时, μ 的单边检验法

未知 σ 时, μ 的单边检验法—例 2.3

- 例 2.1 和例 2.2 中都是检验 $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$, 当 $\bar{X}_n$ 比 μ_0 大许多或小许多时, 都应当否定原假设 H_0, 所以这种检验问题又被称为**双边检验**.
- 下面是**单边检验**问题
- **例 2.3** 在例 2.2 中, 抽查的 9 袋白糖的平均重量 $\bar{X} = 499.412$ 就可以引起我们的怀疑. 这批袋装白糖的平均重量是否不足呢?
- **解** 为了解决这个问题, 我们作原假设

$$H_0 : \ \mu \geq 500 \quad vs \quad H_1 : \mu < 500.$$

- 如果否定了 H_0, 就认定这批袋装白糖的份量不足.
- 由于在 H_0 下, 只知道总体均值 $\mu \geq 500$, 不知道 μ 的具体值, 所以

$$T = \frac{\bar{X}_n - 500}{S/\sqrt{n}}$$

的分布是未知的. 但是这时有

$$T = \frac{\bar{X}_n - 500}{S/\sqrt{n}} \geq T_0 = \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1).$$

- 所以 T 取值较小时应当否定 H_0.
- 因为
$$P\big(T \leq -t_\alpha(n-1)\big) \leq P\big(T_0 \leq -t_\alpha(n-1)\big) = \alpha,$$
- 所以, 当 $T \leq -t_\alpha(n-1)$, 应当否定 H_0, 这时犯第一类错误的概率不超过 α.
- 在本例中, 查表得到 $-t_{0.05}(8) = -1.86$, $T = -2.609 < -1.86$, 所以应当否定 H_0. 认定这批袋装白糖的分量不足时, 犯错误的概率不超过 0.05.
- 例 2.3 中, 以 $\{T \leq -2.609\}$ 为检验的拒绝域时, 刚刚可以拒绝 H_0. 所以检验的 P 值是
$$P = P(T_8 \leq -2.609) = 0.0156.$$
- 分析例 2.1 和 2.3 的问题背景就会看出, 在例 2.1 中应当作双边检验 $H_0 : \mu = 500 \quad vs \quad H_1 : \mu \neq 500$. 因为多装和少装白糖都是不符合生产标准的.
- 在例 2.3 中只需要作单边检验 $H_0 : \mu \geq 500 \quad vs \quad H_1 : \mu < 500$, 因为超市只需要知道袋装白糖不缺斤少两就够了.

未知 σ 时, μ 的单边检验法

- 我们将例 2.3 中的方法总结如下: 如果 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本, σ 未知时,
$$H_0 : \mu \geq \mu_0 \quad vs \quad H_1 : \mu < \mu_0$$
的水平为 α 的拒绝域是
$$\big\{T \leq -t_\alpha(n-1)\big\}, \text{ 其中 } T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}. \tag{2.4}$$
- 如果 $T \leq -t_\alpha(n-1)$ 发生, 就称检验是显著的, 这时否定 H_0 犯错误的概率不超过 α.

- 检验的 P 值是

$$P = P\Big(T_{n-1} \le \frac{\bar{x}_n - \mu_0}{s/\sqrt{n}}\Big), \text{ 其中 } T_{n-1} \sim t(n-1). \tag{2.5}$$

- 同理, 对于来自总体 $N(\mu, \sigma^2)$ 的样本 $X_1, X_2, \cdots, X_n$, σ 未知时, 在检验水平 α 下, 假设

$$H_0 : \mu \le \mu_0 \quad vs \quad H_1 : \mu > \mu_0$$

的拒绝域是

$$\big\{T \ge t_\alpha(n-1)\big\}, \text{ 其中 } T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}. \tag{2.6}$$

- 如果 $T \ge t_\alpha(n-1)$ 发生, 就称检验是显著的, 这时否定 H_0 犯错误的概率不超过 α.

- 检验的 P 值是

$$P = P\Big(T_{n-1} \ge \frac{\bar{x}_n - \mu_0}{s/\sqrt{n}}\Big), \text{ 其中 } T_{n-1} \sim t(n-1).$$

- 以上两种检验方法也称为 t 检验法.

例 2.4

- 糕点厂经理为判断牛奶供应商所供应的鲜牛奶是否被兑水, 对它供应的牛奶进行了随机抽样检查. 测得 12 个鲜牛奶样品的冰点如下,

$$\begin{array}{cccccc} -0.5426 & -0.5467 & -0.5360 & -0.5281 & -0.5444 & -0.5468 \\ -0.5420 & -0.5347 & -0.5468 & -0.5496 & -0.5410 & -0.5405. \end{array}$$

已知天然牛奶的冰点是 -0.545 摄氏度. 问牛奶是否被兑水.

- **解** 设 $n = 12$. 用 X_i 表示第 i 个样品的冰点, 则 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, 参数 μ, σ 未知. 如果牛奶没有被兑水, 则 $\mu = -0.545$.

- 根据测量的数据可以计算出样本均值 $\bar{X}_n = -0.5416$, 样本方差 $S = 0.0061$.

- 由于水的冰点是 0 摄氏度, 所以兑水牛奶的冰点将会提高.

- 现在 $\bar{X}_n > -0.545$, 于是有理由怀疑牛奶被兑水.

- 为了判定牛奶被兑水, 就要看牛奶没被对兑水时, $\bar{X}_n = -0.5416$, $S = 0.0061$ 发生的概率有多大.
- 设 $\mu_0 = -0.545$, 作假设

$$H_0: \ \mu \le \mu_0 \text{ (没兑水)} \quad vs \quad H_1: \mu > \mu_0 \text{ (兑水)}.$$

  如果否定了 H_0, 就判定牛奶被兑水.
- 现在

$$T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} = 1.9308.$$

- 查 $t(11)$ 表得到 $t_{0.05}(11) = 1.796$.
- 由于 $T > 1.796$, 检验是显著的, 所以否定 H_0, 认为牛奶被兑水.
- 判断牛奶被兑水, 犯错误的概率不超过检验水平 $\alpha = 0.05$.
- 本例中检验的 P 值是

$$P = P(T_{11} \ge 1.9308) \approx 0.04.$$

8.2.5 正态近似法

正态近似法

- 设总体 X 分布未知，X 的期望和方差 μ 和 σ^2 未知。希望检验

$$H_0: \mu = \mu_0 \qquad vs \qquad H_1: \mu \ne \mu_0$$

- 对 X 的样本 $X_1, X_2, \ldots, X_n$，如果 n 充分大，则根据 H_0 下

$$Z = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \xrightarrow{\text{d}} \text{N}(0,1)$$

- 可构造水平 α 的否定域

$$W = \{|Z| > z_{\alpha/2}\}$$

- 类似地，对

$$H_0: \mu \le \mu_0 \qquad vs \qquad H_1: \mu > \mu_0$$

  检验的否定域为

$$W = \{Z > z_\alpha\}$$

- 对

$$H_0: \mu \ge \mu_0 \qquad vs \qquad H_1: \mu < \mu_0$$

  检验的否定域为

$$W = \{Z < -z_\alpha\}$$

比例的近似假设检验—双边

- 设总体 $X \sim \mathrm{b}(1,p)$, 样本 $X_1, X_2, \ldots, X_n$，则 $\bar{X}_n$ 是比例 p 的估计，记为 $\hat{p}$。
- 对检验
$$H_0 : p = p_0 \qquad vs \qquad H_1 : p \neq p_0$$
- 在 H_0 成立时
$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$$
- n 充分大时可用否定域
$$W = \{|Z| > z_{\alpha/2}\}$$

比例的近似假设检验—单边

- 对单边检验
$$H_0 : p \leq p_0 \qquad vs \qquad H_1 : p > p_0$$
- 由上面的正态近似法，当 $p = p_0$ 时
$$Z = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$$
- n 充分大时可用否定域
$$W = \{Z > z_\alpha\}$$
- 对左侧单边检验否定域则为
$$W = \{Z < -z_\alpha\}$$

8.3 样本量的选择

功效函数

- 对检验
$$H_0 : \theta \in \Theta_0 \quad vs \qquad H_1 : \theta \in \Theta_1$$
设 W 为 α 水平的检验法，定义
$$P_\theta(W) = \text{真实参数为 } \theta \text{ 时否定 } H_0 \text{ 的概率}$$
为检验法 W 的功效函数。

- 当 H_0 成立时，$P_\theta(W)$ 是第一类错误概率。
- 当 H_1 成立是，$P_\theta(W)$ 是 1 减去第二类错误概率，称为检验的功效。
- 检验法控制 $P_\theta(W) \leq \alpha, \forall \theta \in \Theta_0$。
- 给定水平后功效越高检验法越好，但是在 Θ_0 和 Θ_1 交界的地方功效可以只有 α，即第二类错误概率可以很高。
- 在 Θ_0 和 Θ_1 交界的地方第二类错误概率大是可以容忍的。

功效函数的例子

- 考虑正态总体 σ^2 已知时检验

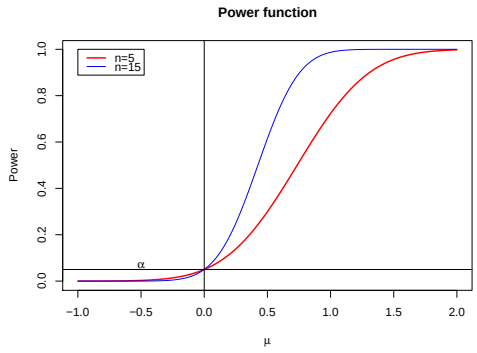
$$H_0 : \mu \leq \mu_0 \quad vs \quad H_1 : \mu > \mu_0$$

- 否定域为

$$W = \left\{ \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \right\}$$

- 功效函数

$$P_\mu(W) = P(Z + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} > z_\alpha)$$



8.4 均值比较的检验

独立两总体比较—抽样分布

- 设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$.
- $X_1, X_2, \cdots, X_n$ 是来自 X 的样本, $Y_1, Y_2, \cdots, Y_m$ 是来自 Y 的样本, 考虑有关 μ_1 和 μ_2 的比较的假设检验问题.

- 以下设总体 X 和总体 Y 独立, 于是

$$X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_m$$

相互独立.

- 用 $\bar{X}_n$, $\bar{Y}_m$ 分别表示 X, Y 的样本均值, 用 S_1^2, S_2^2 分别表示 X, Y 的样本方差, 由 §7.4 (4.12) 知

$$Z = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \sim N(0,1). \tag{4.1}$$

- 当未知 σ_1^2, σ_2^2, 但已知 $\sigma_1^2 = \sigma_2^2$ 时, 设 $\sigma^2 = \sigma_1^2$, 由 §7.4 的 (4.16) 得

$$T_0 = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_1 - \mu_2)}{S_W\sqrt{1/n + 1/m}} \sim t(n+m-2). \tag{4.4}$$

其中

$$S_W^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}, \tag{4.3}$$

8.4.1 已知 σ_1^2 , σ_2^2 时, μ_1, μ_2 的检验

已知 σ_1^2 , σ_2^2 时, μ_1, μ_2 的检验

- **例 4.1** 甲乙两公司都生产 700MB(兆字节) 的光盘, 从甲的产品中抽查了 7 张光盘, 从乙生产的产品中抽查了 9 张光盘. 分别测得他们的储量如下

X(甲)	683.7 682.5 683.5 678.7 681.1 680.80 677.9
Y(乙)	681.5 682.7 674.2 674.6 680.7 677.8 681.0
	681.4 681.1

现在已知甲的光盘储量 $X \sim N(\mu_1, 2)$, 乙的光盘储量 $Y \sim N(\mu_2, 3)$. 在显著性水平 $\alpha = 0.05$ 下, 甲乙两家光盘的平均储量有无显著性差异? 计算检验的 P 值.

- **解** $n = 7$, $m = 9$, $\sigma_1^2 = 2$, $\sigma_2^2 = 3$. 作假设

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2. \tag{4.5}$$

- 设 Z 由 (4.1) 定义, 在 H_0 下,

$$\xi = \frac{(\bar{X}_n - \bar{Y}_m)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \sim N(0,1). \tag{4.6}$$

- $P(|\xi| \geq 1.96) = 0.05$.
- H_0 的水平 0.05 拒绝域是

$$W = \big\{\, |\xi| \geq 1.96 \,\big\}. \tag{4.7}$$

- 经过计算得到 $\bar{X}_n = 681.17$, $\bar{Y}_m = 679.44$,

$$\xi = \frac{681.17 - 679.44}{\sqrt{2/7 + 3/9}} = 2.1988 > 1.96.$$

  可以在水平 0.05 下认为这两家光盘的平均储量有显著差异.

- 检验的 P 值是 $P = P(|Z| \geq 2.1988) = 2P(Z > 2.1988) = 0.0279$.

例 4.2

- 在例 4.1 中, 能否在检验水平 0.02 下认为甲光盘的平均储量大于乙光盘的平均储量? 计算检验的 P 值.
- **解** 由于 $\bar{X}_n = 681.17 > \bar{Y}_m = 679.44$, 所以初看起来 $\mu_1 > \mu_2$.
- 作假设

$$H_0 : \mu_1 \leq \mu_2 \quad vs \quad H_1 : \mu_1 > \mu_2. \tag{4.8}$$

- 在 H_0 下, $\bar{X}_n - \bar{Y}_m$ 应当比较小, ξ 取值较大时应当否定 H_0.
- 查表 C1(续) 知道 $z_{0.02} = 2.054$.
- 所以 H_0 的水平 0.02 拒绝域是 $W = \{\xi > 2.054\}$.
- 现在 $\xi = 2.1988 > 2.054$, 所以可以在水平 0.02 下认为甲光盘的平均储量大于乙光盘的平均储量.
- 检验的 P 值是 $P = P(Z > 2.1988) = 0.0139$.
- 在例 4.2 中, 由于 P 值小于 0.02, 所以在水平 0.05 下, 更能否定 $H_0 : \mu_1 \leq \mu_2$.
- 分析例 4.1 和例 4.2 看出: 由于 $z_{\alpha/2} > z_\alpha$, 所以对双边假设能够否定 $H_0 : \mu_1 = \mu_2$ 时, 在相同的检验水平下, 只要 $\bar{X}_n > \bar{Y}_m$, 对单边假设更能够否定 $H_0 : \mu_1 < \mu_2$.

8.4.2 未知 σ_1^2, σ_2^2, 但已知 $\sigma_1^2 = \sigma_2^2$ 时, $\mu_1 - \mu_2$ 的检验

未知 σ_1^2, σ_2^2, 但已知 $\sigma_1^2 = \sigma_2^2$ 时, $\mu_1 - \mu_2$ 的检验

- **例 4.3** 甲乙两公司用同型号的组装线分别生产自己的 128MB(兆字节) 闪盘, 从甲的产品中抽查了 7 只, 从乙生产的产品中抽查了 8 只闪盘. 分别测得他们的储量如下

X(甲)	125.5	124.3	126.12	126.2	124.8	127.2	127.19
Y(乙)	124.7	125.0	124.8	124.5	125.4	124.1	126.9	124.6

  设甲的闪盘储量和乙的闪盘储量都服从正态分布, 且有相同的标准差.

-

  (1) 在显著性水平 $\alpha = 0.05$ 下, 这两家光盘的平均储量有无显著差异,

  (2) 在显著性水平 $\alpha = 0.1$ 下, 这两家光盘的平均储量有无显著差异,

  (3) 在显著性水平 $\alpha = 0.05$ 下, 能否认为 $\mathrm{E}X > \mathrm{E}Y$.

- **解** 设 $\mu_1 = \mathrm{E}X$, $\mu_2 = \mathrm{E}Y$, $n = 7$, $m = 8$.

- (1) 作假设

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2. \tag{4.9}$$

- 在 H_0 下, 由 (4.4) 知道

$$T = \frac{\bar{X}_n - \bar{Y}_m}{S_W \sqrt{1/n + 1/m}} \sim t(n + m - 2), \tag{4.10}$$

- 于是 $P(|T| \geq t_{0.05/2}(13)) = 0.05$.

- 查 t 分布表得到 $t_{0.05/2}(13) = 2.16$, 于是 H_0 的水平 0.05 拒绝域是

$$W = \{|T| \geq 2.16\}. \tag{4.11}$$

- 经过计算得到

$$\begin{aligned} &\bar{X}_n = 125.9, \ \ \bar{Y}_m = 125.0, \\ &S_1^2 = 1.1122, \ \ S_2^2 = 0.8552. \\ &S_W^2 = 0.9648, \end{aligned}$$

- 最后得到
$$T = \frac{125.9 - 125.0}{S_W\sqrt{1/7 + 1/8}} = 1.805 < 2.16.$$
所以不能在水平 0.05 下认为这两家闪盘的平均储量有显著性差异.

- (2) 由于 n, m 都较小, T 的值又比较大, 我们还不情愿接受 H_0.

- 查 t 分布表, 得到 $t_{0.1/2}(13) = t_{0.05}(13) = 1.771$. 根据 (4.10), 在 H_0 下
$$P(|T| > 1.771) = 0.1.$$

- 现在 $T = 1.805 > 1.771$, 所以我们可以在水平 0.1 下否定 H_0, 认为这两家闪盘的平均储量有显著性差异.

- (3) 因为 $\bar{X}_n = 125.9 > \bar{Y}_m = 125.0$, 所以可能有 $\mu_1 > \mu_2$.

- 作假设
$$H_0 : \mu_1 \le \mu_2 \quad vs \quad H_1 : \mu_1 > \mu_2.$$

- 设 T_0, T 分别在 (4.4) 和 (4.10) 中定义.

- 在 H_0 下, $(\mu_1 - \mu_2) \le 0$, 故
$$T \le T_0 = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_1 - \mu_2)}{S_W\sqrt{1/n + 1/m}} \sim t(n + m - 2),$$
$$P(T \ge t_{0.05}(13)) \le P(T_0 \ge t_{0.05}(13)) = 0.05.$$

- 由于已得到 $t_{0.05}(13) = 1.771$. 于是得到 H_0 的水平 0.05 拒绝域是
$$W = \{T \ge 1.771\}.$$

- 现在 $T = 1.805 > 1.771$, 所以我们可以在水平 0.05 下否定 H_0. 认为甲光盘的平均储量比乙光盘的平均储量大.

- 从例 4.3 看出, 在相同的显箸性水平下, 和双边检验比较, 单边检验更易于得出否定 H_0 的结果.

8.4.3 成对数据的假设检验

成对数据的假设检验

- **例 4.4** 在考古学中，人们可以用碳 14 方法确定发掘物的年代.

- 这是由于不论何种动植物生长在何处，由于新陈代谢的原因, 存活时其细胞组织中每一克碳内所含的碳 14 数目是相同的.
- 但是当动植物的生命停止后, 得不到补充的碳 14 衰变时能放出 β 粒子，其半衰期约为 5730 年. 即大约经过 5730 年下降一半，经 11460 年减少到 1/4 等. 依此类推，就能根据动植物残骸中碳 14 的含量来确定其停止呼吸的时间.
- 现在考古学家们在某建筑工地陆续发掘出了已经碳化了的谷物种子的 12 个样品, X,Y 两个考古单位分别对这 12 个样品用碳 14 方法进行了年代测定 (单位: 万年), 结果如下.

X	0.81	0.57	0.69	0.68	0.53	0.72
Y	0.72	0.63	0.53	0.70	0.69	0.80
X	0.59	0.84	0.61	0.75	0.72	0.60
Y	0.69	0.57	0.67	0.53	0.63	0.63

- (1) 在检验水平 0.05 下, 这两家的测量年代有无明显的差异?
- (2) 认为有明显差异时犯错误的概率是多少?
- **解** 用 $X_1, X_2, \cdots, X_2$ 分别表示甲单位对第 $1, 2, \cdots, 12$ 号样品的测定, 用 $Y_1, Y_2, \cdots, Y_2$ 分别表示乙单位对第 $1, 2, \cdots, 12$ 号样品的测定.
- 引入

$$Z_1 = X_1 - Y_1,\ Z_2 = X_2 - Y_2,\ \cdots,\ Z_n = X_n - Y_n.$$

- 设 $Z = X - Y \sim \mathrm{N}(\mu, \sigma^2)$, $Z_1, Z_2, \ldots, Z_n$ 是 Z 的样本。
- 我们要检验的是假设

$$H_0 : \mu = 0 \quad vs \quad H_1 : \mu \neq 0.$$

- 问题已经转化成一个样本的 t 检验问题.
- H_0 的水平 0.05 拒绝域是

$$W = \{|T| \geq t_{0.05/2}(11)\},\ T = \frac{\overline{Z}_n}{S_Z/\sqrt{n}}.$$

- 由于 $t_{0.05/2}(11) = 2.201$,

$$T = \frac{\overline{Z}_n}{S_z/\sqrt{n}} = 0.6766 < 2.201$$

所以不能否定 H_0, 即不能认为两单位的测定有明显的差异.

- (2) 由于检验的 P 值为

$$P = P(|T_{11}| > 0.6766) = 2P(T_{11} > 0.6766) = 0.512,$$

说明否定 H_0 的证据不足。

- 本例中, 如果采用方差相等的检验法（当然，理论上不合适）, 也不能否定 H_0.

8.4.4 未知 σ_1^2, σ_2^2 时, μ_1, μ_2 的检验

未知 σ_1^2, σ_2^2 时, μ_1, μ_2 的检验

- 未知 σ_1^2, σ_2^2 时, 对

$$\eta_n = \frac{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}{\sqrt{S_1^2/n + S_2^2/m}} = \frac{\sqrt{\sigma_1^2 + \sigma_2^2(n/m)}}{\sqrt{S_1^2 + S_2^2(n/m)}},$$

利用 $S_1^2 \to \sigma_1^2, S_2^2 \to \sigma_2^2$, wp1., 得到当 $n/m \to 1$,

$$\lim_{n,m\to\infty} \eta_n \to 1, \text{ wp1.}$$

- 于是利用 §7.3 的定理 3.9, 在 H_0 下得到, 当 $n/m \to 1$, $n \to \infty$ 时,

$$T = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_1^2/n + S_2^2/m}} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}\eta_n = Z\eta_n \to^d N(0,1).$$

- 即当样本量 n, m 都较大时, T 近似服从 $N(0,1)$ 分布, 从而可以得到 $H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$ 的检验法

$$W = \{|T| > z_{\alpha/2}\}, \text{ 其中 } T = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_1^2/n + S_2^2/m}}.$$

例 4.5

- X, Y 两个渔场在初春放养相同的鳜鱼苗, 但是采用不同的方法喂养. 入冬时, 从第一渔场打捞出 59 条鳜鱼, 从第二渔场打捞出 41 条鳜鱼. 分别秤出他们的平均重量和样本标准差如下 (单位:kg),

$$\bar{X}_n = 0.59,\ S_1 = 0.2,\ \ \bar{Y}_m = 0.62,\ S_2 = 0.21.$$

- (1) 在显著性水平 0.05 下, 就鳜鱼的平均重量来讲, 两个渔场的养殖结果有无显著差异;

- (2) 计算检验的 P 值.

- **解** 对 $n=59, m=41$, 容易计算出

$$Z=\frac{\bar{X}_n-\bar{Y}_m}{\sqrt{S_1^2/n+S_2^2/m}}=-0.7164$$

- 由于 $|Z|<1.96$, 所以在水平 0.05 下不能否定 $H_0: \mathrm{E}X=\mathrm{E}Y$. 不能认为两个渔场的养殖结果有显著性差异.

- 检验的 P 值是

$$P=P(|Z|>0.7164)=2[1-\Phi(0.7164)]\approx 0.47.$$

否定 H_0 的证据不足。

8.5 方差的假设检验

方差的假设检验

- **例 5.1** 在例 4.5 中, 从第一渔场打捞出 59 条鳜鱼, 从第二渔场打捞出 41 条鳜鱼. 分别秤出他们的样本标准差如下 (单位:kg)

$$S_1=0.2,\ S_2=0.21.$$

对 $\sigma_0^2=0.18^2$, 在显著性水平 $\alpha=0.05$ 下, 解决以下检验问题.

- (1) $H_0: \sigma_1^2=\sigma_0^2 \quad vs \quad H_1: \sigma_1^2\neq\sigma_0^2$,

- (2) $H_0: \sigma_1^2\leq\sigma_0^2 \quad vs \quad H_1: \sigma_1^2>\sigma_0^2$,

- (3) $H_0: \sigma_1^2\geq\sigma_2^2 \quad vs \quad H_1: \sigma_1^2<\sigma_2^2$.

- **解** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(\mu_1, \sigma_1^2)$ 的样本, 则

$$\xi_1=\frac{(n-1)S_1^2}{\sigma_1^2}\ \sim\ \chi^2(n-1),\ n=59.$$

- (1) 在 H_0 下

$$\xi=\frac{(n-1)S_1^2}{\sigma_0^2}\ \sim\ \chi^2(n-1),\ n=59.$$

- S_1^2 是 σ_0^2 的强相合估计, 所以 ξ 取值过大和过小都是拒绝 H_0 的依据.

- 用 $\chi_\alpha^2(n-1)$ 表示 $\chi^2(n-1)$ 的上 α 分位数, 则可以构造出假设 (1) 的水平 α 拒绝域

$$W_1=\{\xi\leq\chi_{1-\alpha/2}^2(n-1)\}\bigcup\{\xi\geq\chi_{\alpha/2}^2(n-1)\}. \tag{5.1}$$

- 这是因为在 H_0 下, 有

$$P(W_1) = P(\xi \le \chi^2_{1-\alpha/2}(n-1)) + P(\xi \ge \chi^2_{\alpha/2}) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$

- 本例中, 查表得到 $\chi^2_{1-\alpha/2}(58) = \chi^2_{0.975}(58) = 38.84$, $\chi^2_{\alpha/2}(58) = 80.94$, 否定域是 $W_1 = \{\xi \le 38.84\} \bigcup \{\xi \ge 80.94\}$. 现在

$$\xi = \frac{58S_1^2}{0.18^2} = 71.60 \notin W_1.$$

  所以在检验水平 0.05 下不能否定 H_0.

- 本检验是用 χ^2 分布完成的, 所以又称为χ^2 **检验**.

- (2) 在 $H_0 : \sigma_1^2 \le \sigma_0^2$ 下, σ_1^2 是真参数,

$$\xi = \frac{(n-1)S_1^2}{\sigma_0^2} \le \frac{(n-1)S_1^2}{\sigma_1^2} = \xi_1 \sim \chi^2(n-1).$$

- 对 $\sigma_1 \le \sigma_0$,

$$P_{\sigma_1}(\xi > \lambda) \le P_{\sigma_1}(\xi_1 > \lambda)$$

- 取 $\lambda = \chi^2_\alpha(n-1)$, 则

$$P_{\sigma_1}(\xi > \lambda) \le \alpha$$

- ξ 取值过大是拒绝 H_0 的依据, 于是得到 H_0 的水平 0.05 拒绝域

$$W_2 = \{\xi \ge \chi^2_{0.05}(58)\} = \{\xi \ge 76.78\}$$

- 现在 $\xi = 71.6$, 仍然不能否定 H_0.

- 这个检验也是用 χ^2 分布完成的, 也称为 χ^2 检验.

- (3) 在 H_0 下, σ_1^2/σ_2^2 大于 1, 所以 $F \equiv S_1^2/S_2^2$ 取值也应当较大, F 较小时应当拒绝 H_0.

- 由于总体 X 和 Y 独立, 所以利用 §7.3 的定理 3.7 知道

$$F = \frac{S_1^2}{S_2^2} \ge \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1).$$

- 于是得到 H_0 的水平 α 否定域

$$W_3 = \{F \le F_{1-\alpha}(58, 40)\}.$$

  经过查表和计算得到 $F_{1-0.05}(58, 40) = 1/F_{0.05}(40, 58) = 0.625$.

$$F = 0.2^2/0.21^2 = 0.907 \ge 0.625.$$

  所以不能在显著性水平 0.05 下否定 H_0.

- 我们还不能根据数据判定 X 的方差小于 Y 的方差.
- 这里的检验是用 F 分布进行的, 所以又称为 F 检验.

8.6 比例的假设检验

8.6.1 小样本情况下的假设检验

小样本单个比例的右侧检验

- 设总体 $X \sim \mathrm{B}(1,p)$, $X_1, X_2, \ldots, X_n$ 是 X 的样本，$p_0 \in (0,1)$ 为已知常数。
- 单样本单边比例检验问题

$$H_0:\ p \le p_0 \longleftrightarrow H_1:\ p > p_0 \tag{6.1}$$

- 用统计量

$$\xi_n = X_1 + X_2 + \cdots + X_n \sim \mathrm{B}(n,p),$$

取否定域为

$$W_1 = \{\xi_n \ge B_\alpha(n,p_0)\} \tag{6.2}$$

其中 $B_\alpha(n,p_0)$ 是 $\mathrm{B}(n,p_0)$ 的上侧 α 分位数。

- 当 H_0 成立即 $p \le p_0$ 时，设 η 为 $\mathrm{B}(n,p_0)$ 随机变量，则

$$P(W_1) = P(\xi_n \ge B_\alpha(n,p_0)) \le P(\eta \ge B_\alpha(n,p_0)) \le \alpha$$

- 于是，对检验问题 (6.1)，给定检验水平 α，设 n 个独立样本点中成功个数为 ξ_n，查二项分布上侧分位数表得 $B_\alpha(n,p_0)$，当 $\xi_n \ge B_\alpha(n,p_0)$ 拒绝 H_0，认为在检验水平 α 下成功概率 p 显著地高于 p_0; 否则，认为在检验水平 α 下成功概率 p 不显著高于 p_0。

小样本单个比例的左侧检验

- 类似地，对检验问题

$$H_0:\ p \ge p_0 \longleftrightarrow H_1:\ p < p_0 \tag{6.3}$$

取拒绝域为

$$W_2 = \{\xi_n \le n - B_\alpha(n,1-p_0)\} \tag{6.4}$$

- 记 $q=1-p$, $q_0=1-p_0$, 这时 $\zeta_n=n-\xi_n\sim \mathrm{B}(n,q)$, 令 $\tilde{\eta}\sim \mathrm{B}(n,q_0)$, 则 H_0 成立即 $p\geq p_0$ 时 $q\leq q_0$，有

$$\begin{aligned} P(W_2) =& P(\xi_n\leq n-B_\alpha(n,q_0))=P(\zeta_n\geq B_\alpha(n,q_0)) \\ \leq& P(\tilde{\eta}\geq B_\alpha(n,q_0))\leq \alpha. \end{aligned}$$

- 给定检验水平 α 后，查二项分布上侧分位数表得 $B_\alpha(n,1-p_0)$, 当且仅当 $\xi_n\leq n-B_\alpha(n,1-p_0)$ 时拒绝 H_0。

小样本单个比例的双侧检验

- 对检验问题

$$H_0:\ p=p_0 \longleftrightarrow H_1:\ p\neq p_0 \tag{6.5}$$

- 取否定域为

$$W_3=\{\xi_n\leq n-B_{\frac{\alpha}{2}}(n,1-p_0)\text{或}\xi_n\geq B_{\frac{\alpha}{2}}(n,p_0)\} \tag{6.6}$$

二项分布上侧分位数定义

- 设随机变量 $X\sim \mathrm{B}(n,p)$。
- 若非负整数 $\lambda\in\{0,1,\ldots,n\}$ 使得

$$P(X\geq\lambda)\leq\alpha, P(X\geq\lambda-1)>\alpha \tag{*}$$

  则称 λ 为 $\mathrm{B}(n,p)$ 分布的上侧 α 分位数。记作 $B_\alpha(n,p)$。

- 实际上，因为

$$P(X\geq m)=\sum_{k=m}^{n}C_n^k p^k(1-p)^{n-k},\ m=0,1,\ldots,n$$

  是一个严格单调递减序列，所以对任意 $\alpha\in(0,1)$ 一定能找到唯一的 λ 满足 (*) 式。

- 简单而言可以用穷举法找到 $B_\alpha(n,p)$ 的值。

例 6.1

- 小区物业称业主满意率 $p \geq 80\%$。
- 随机抽样调查 50 户，有 33 户回答满意，$\hat{p} = 30/50 = 66\%$。
- 能否认为满意率真的 $p \geq 80\%$?
- **解答:** 如何选择单侧检验的方向？
- 数据中实际满意率是低于 $p_0 = 0.8$ 的，所以对立假设设为 $p < 0.8$，这样零假设用 $p \geq 0.8$。
- 设 ξ_n 是回答满意的户数，否定域为

$$\xi_n \leq n - B_\alpha(n, p_0).$$

- 取 $\alpha = 0.05$，$n = 50$, $\xi_n = 33$, 查表得 $B_{0.05}(50, 0.8) = 45$。现在 $\xi_n = 33 \leq 45$ 成立，拒绝 H_0，在 0.05 水平下业主满意率显著地低于 80%。

8.6.2 大样本情况下单个比例的假设检验

大样本情况下单个比例的假设检验

- 对单个比例 p，设 ξ_n 是 n 个独立抽样中成功的个数，则 $\xi_n \sim \mathrm{B}(n, p)$, 当 n 很大时（一般要求成功和失败个数都超过 5 个），根据中心极限定理，ξ_n 近似服从正态分布 $\mathrm{N}(np, np(1-p))$。
- 令 $\hat{p} = \xi_n/n$ 为样本中成功比例。则

$$\eta = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

  近似服从标准正态分布。
- 当 n 很大时另一近似为

$$Z = \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}}$$

  近似服从标准正态分布。

双侧检验

- 对双侧问题

$$H_0:\ p = p_0 \longleftrightarrow H_1:\ p \neq p_0$$

在 H_0 下

$$\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

近似服从标准正态分布。

- 给定检验水平 α，取否定域为

$$W = \left\{ \frac{|\hat{p} - p_0|}{\sqrt{p_0(1-p_0)/n}} \geq z_{\frac{\alpha}{2}} \right\} \tag{6.7}$$

右侧检验

- 对右侧检验问题

$$H_0:\ p \leq p_0 \longleftrightarrow H_1:\ p > p_0$$

- 取否定域为

$$W = \left\{ \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \geq z_{\alpha} \right\} \tag{6.8}$$

左侧检验

- 对左侧检验问题

$$H_0:\ p \geq p_0 \longleftrightarrow H_1:\ p < p_0$$

- 取否定域为

$$W = \left\{ \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq -z_{\alpha} \right\} \tag{6.9}$$

- 这样的比例检验方法称为**正态逼近法**。

例 6.2

- 收藏家一年中购入了 98 幅名画，经鉴定其中 26 幅是赝品。能否认为该收藏家的鉴定准确率大于等于 75%?
- **解答：** 准确率 p 的估计为 $\hat{p} = (98 - 26)/98 = 0.7347$，低于 75%，所以取单侧检验的方向为

$$H_0 : p \geq 0.75 \longleftrightarrow H_1 : p < 0.75$$

- 用正态近似法，取检验水平 $\alpha = 0.05$, $n = 98$, 否定域为

$$\left\{ \frac{\hat{p} - 0.75}{\sqrt{\hat{p}(1-\hat{p})/n}} < -1.645 \right\}$$

- 计算得

$$\frac{\hat{p} - 0.75}{\sqrt{\hat{p}(1-\hat{p})/n}} = -0.3432$$

  未落入否定域，在 0.05 水平下不能拒绝鉴定准确率大于等于 75% 的假设。

- 注意，如果我们去假设的方向为右侧问题

$$H_0 : p \leq 0.75 \longleftrightarrow H_1 : p > 0.75$$

- 0.05 水平否定域为

$$\left\{ \frac{\hat{p} - 0.75}{\sqrt{\hat{p}(1-\hat{p})/n}} > 1.645 \right\}$$

- 现在统计量

$$\frac{\hat{p} - 0.75}{\sqrt{\hat{p}(1-\hat{p})/n}} = -0.3432$$

  也没有落入否定域，在 0.05 水平下不能拒绝鉴定准确率小于等于 75% 的假设。

- 这是通常的假设检验方法的局限性：当不能拒绝 H_0 时，最后的结论往往是不确定的。

8.6.3 大样本情况下两个总体比例的比较

大样本情况下两个总体比例的比较

- 设总体 $X \sim B(1, p_1)$, 总体 $Y \sim B(1, p_2)$ 与 X 独立。
- 设 $X_1, \ldots, X_n$ 为来自 X 的样本，$S_1 = X_1 + \cdots + X_n$ 为 n 个独立抽样的成功次数，$\hat{p}_1 = S_1/n$ 为成功比例；设 $Y_1, \ldots, Y_m$ 为来自 Y 的样本，$S_2 = Y_1 + \cdots + Y_m$ 为 m 个独立抽样的成功次数，$\hat{p}_2 = S_2/m$ 为成功比例。
- 由中心极限定理，当 n, m 都很大时近似有

$$\begin{aligned}\hat{p}_1 \sim & \mathrm{N}(p_1, p_1(1-p_1)/n),\\ \hat{p}_2 \sim & \mathrm{N}(p_2, p_2(1-p_2)/m),\end{aligned}$$

- 从而近似有

$$\hat{p}_1 - \hat{p}_2 \sim \mathrm{N}\left(p_1 - p_2, \frac{1}{n}p_1(1-p_1) + \frac{1}{m}p_2(1-p_2)\right)$$

- 实际使用时，设两个样本中的成功和失败个数都在 5 个以上。

右侧检验

- 对右侧检验问题

$$H_0 : \ p_1 \le p_2 \longleftrightarrow H_1 : \ p_1 > p_2$$

- 取统计量

$$\xi = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{1}{n}\hat{p}_1(1-\hat{p}_1) + \frac{1}{m}\hat{p}_2(1-\hat{p}_2)}}$$

  当 $p_1 = p_2$ 时且 n, m 都很大时 ξ 近似标准正态分布。
- 取水平 α 的否定域为

$$\{\xi \ge z_\alpha\}$$

左侧检验

- 对左侧检验问题

$$H_0 : \ p_1 \ge p_2 \longleftrightarrow H_1 : \ p_1 < p_2$$

- 取水平 α 的否定域为

$$\{\xi \le -z_\alpha\}$$

双侧检验

- 对双侧检验问题

$$H_0:\ p_1 = p_2 \longleftrightarrow H_1:\ p_1 \neq p_2$$

- 在 H_0 下设 $p_1 = p_2 = p_0$，则近似地有

$$\hat{p}_1 - \hat{p}_2 \sim \mathrm{N}\left(0, \left[\frac{1}{n} + \frac{1}{m}\right] p_0(1-p_0)\right)$$

  其中 p_0 可以用 $\hat{p}_0 = (S_1 + S_2)/(n+m)$ 来估计。

- 取统计量

$$\eta = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\hat{p}_0(1-\hat{p}_0)}}$$

  当 $p_1 = p_2$ 时且 n, m 都很大时 η 近似标准正态分布。

- 取水平 α 的否定域为

$$\{|\eta| \geq z_{\frac{\alpha}{2}}\}$$

- 以上的三个两总体比例比较方法也称为**正态逼近法**。

例 6.3

- $n = 1230$ 位男应届毕业生和 $m = 1542$ 名女应届毕业生参加由政府组织的就业洽谈会。结果有 251 名男生和 232 名女生求职成功。

- 假设除性别不同外学生表现没有其它方面的系统差异，问此次招聘有无性别歧视。

- *解答*： 设男生成功率为 p_1，女生成功率为 p_2，问题是 $p_1 = p_2$ 是否成立。

- 检验问题是

$$H_0: p_1 = p_2 \longleftrightarrow H_1: p_1 \neq p_2$$

- 样本量足够大，可以用正态逼近法。

- 水平 0.05 否定域为 $\{|\eta| > 1.96\}$。

- 现在 $n = 1230, m = 1542$, $\hat{p}_1 = 251/1230 = 0.2041, \hat{p}_2 = 232/1542 = 0.1505$, $\hat{p}_0 = (251 + 232)/(1230 + 1542) = 0.1742$。
- 计算统计量值

$$\begin{aligned}\eta =& \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\hat{p}_0(1 - \hat{p}_0)}} \\ =& \frac{0.2041 - 0.1505}{\sqrt{\left(\frac{1}{1230} + \frac{1}{1542}\right)0.1742(1 - 0.1742)}} = 3.697 > 1.96\end{aligned}$$

- 在 0.05 水平下否定 H_0，认为招聘男女比例有显著差异，可以认为有性别歧视，而且是不利于女生。

例 6.4

- 在第六章例 5.3 中，被实验组的 20 万儿童注射疫苗，发病率为 $\hat{p}_1 = 28 \times 10^{-5}$, 对照组的 20 万儿童发病率为 $\hat{p}_2 = 71 \times 10^{-5}$。
- 问：疫苗是否有效？
- 因为做出疫苗有效的结论是需要很慎重的，所以我们取双侧检验 (这是医学中通常的做法)

$$H_0 : p_1 = p_2 \longleftrightarrow H_1 : p_1 \neq p_2$$

- 样本足够大。
- 取检验水平 0.01，否定域为 $\{|\eta| > 2.58\}$。
- $\hat{p}_0 = (200000 \times 0.00028 + 200000 \times 0.00071)/400000 = 0.000495$。统计量为

$$\begin{aligned}\eta =& \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\hat{p}_0(1 - \hat{p}_0)}} \\ =& \frac{0.00028 - 0.00071}{\sqrt{\left(\frac{1}{200000} + \frac{1}{200000}\right)0.000495(1 - 0.000495)}} = -6.311\end{aligned}$$

- $|\eta| > 2.58$, 在 0.01 水平下拒绝零假设，认为处理组发病率与对照组发病率有显著差异，疫苗有效。
- 可进一步计算检验 p 值为

$$P(|Z| > |-6.311|) = 2[1 - \Phi(6.311)] = 2.77 \times 10^{-10}$$

所以显著性差异的证据是很强的。

8.7 总体分布的假设检验

总体分布的假设检验

- 设 $X_1, X_2, \ldots, X_n$ 是总体 X 的样本，设 X 的分布函数 $F(x)$ 未知，$F_0(x)$ 是一个已知的分布函数。
- 考虑假设检验问题

$$H_0: F(x) \equiv F_0(x) \longleftrightarrow H_1: F(x) \not\equiv F_0(x)$$

- 这样的检验问题称为拟合优度检验。

离散分布情形

- 设 F_0 代表了一个离散分布，设 X 仅在 $\{a_1, a_2, \ldots, a_m\}$ 中取值，在 F_0 下 $P(X = a_j) = p_j$, $j = 1, 2, \ldots, m$。
- 设 n 个样本点中 a_j 出现的次数为 f_j, 称为频数，当 H_0 成立时，f_j 应该与 np_j 近似相等。
- 取统计量

$$\xi = \sum_{j=1}^{m} \frac{(f_j - np_j)^2}{np_j}$$

  当 n 充分大时 ξ 在 H_0 下近似服从 $\chi^2(m-1)$ 分布。
- 称 ξ 为拟合优度卡方统计量。
- 取水平 α 的拒绝域为

$$\{\xi \geq \chi^2_\alpha(m-1)\}$$

推广 1

- 如果 F_0 是一类分布，其中还包含 k 个未知参数，可以用最大似然估计方法求得参数估计，用估计的参数计算 F_0 下 $p_j = P(X = a_j)$ 的值。
- 仍计算统计量 ξ，但取否定域为

$$\{\xi \geq \chi^2_\alpha(m-k-1)\}$$

  (卡方临界值的自由度等于组数减一再减未知参数个数)

推广 2

- 当 F_0 代表的离散分布有无穷多个取值，或者取值个数有限但是有些组频数太小时（一般要求频数超过 5），可以把频数很小的类合并。
- 仍记 m 为合并后的组数，这时 p_j 为合并过后的第 j 组在 H_0 下的理论概率，f_j 是合并过后第 j 组的频数。
- 仍有

$$\xi = \sum_{j=1}^{m} \frac{(f_j - np_j)^2}{np_j}$$

- 水平 α 的拒绝域为

$$\{\xi \geq \chi_\alpha^2(m-k-1)\}$$

其中 k 是 F_0 中未知参数的个数。

推广 3—连续分布情形

- 如果 F_0 是一个连续型分布，适当选择分点 $b_0, b_1, \ldots, b_m$，构成区间

$$(b_0, b_1], (b_1, b_2], \ldots, (b_{m-1}, b_m]$$

把 X 的取值空间分成 m 个组，设 f_j 为第 j 组的频数，令 $p_j = P(X \in (b_{j-1}, b_j]|H_0) = F_0(b_j) - F_0(b_{j-1})$。

- 如果 F_0 中有未知参数，计算 p_j 时先获得参数的最大似然估计，$F_0(\cdot)$ 使用参数最大似然估计计算。
- 仍计算统计量

$$\xi = \sum_{j=1}^{m} \frac{(f_j - np_j)^2}{np_j}$$

- 否定域为

$$\{\xi \geq \chi_\alpha^2(m-k-1)\}$$

其中 k 是 F_0 中未知参数的个数。

例 7.1

- 在第二章例 2.2 中，1500 年至 1931 年的 $n = 432$ 年中，全世界发生了 299 次比较重要的战争。
- 用 X_j 表示第 j 年发生的战争次数，认为 $X_1, \ldots, X_n$ 是来自泊松分布的样本。
- 估计泊松参数为 $\hat{\lambda} = \sum X_j/n = 0.69$, 合并比较小的概率，把总体 X 的取值分为

$$0, 1, 2, [3, \infty)$$

- 在 Poisson(0.69) 下计算 $m = 4$ 个组分别的概率

$$\begin{aligned} p_1 =& P(X = 0) = 0.502,\ p_2 = P(X = 1) = 0.346, \\ p_3 =& P(X = 2) = 0.119,\ p_4 = P(X \geq 3) = 1 - p_1 - p_2 - p_3 = 0.033 \end{aligned}$$

- 各组的频数为 $(223, 142, 48, 19)$。
- 检验统计量为

$$\begin{aligned} \xi =& \frac{(223 - 432 \times 0.502)^2}{432 \times 0.502} + \frac{(142 - 432 \times 0.346)^2}{432 \times 0.346} \\ &+ \frac{(48 - 432 \times 0.119)^2}{432 \times 0.119} + \frac{(19 - 432 \times 0.033)^2}{432 \times 0.033} \\ =& 2.346 \end{aligned}$$

- 水平 0.05 的否定域为 $\{\xi \geq \chi^2_{0.05}(4 - 1 - 1)\} = \{\xi \geq 5.991\}$, $\xi = 2.346$ 未落入否定域，所以在 0.05 水平下可以认为样本来自泊松分布。
- 注意拟合优度检验承认 H_0 时可以承认总体服从 F_0 分布，但是和其它检验承认 H_0 的问题类似，如果 H_0 换成某个其它的已知分布 G_0，也可能会获得承认。

例 7.2

- 在第六章例 3.1 中，列出了某公共图书馆在一年中通过随机抽样调查得到了 60 天中每天的读者借书数, 能否认为这批数据是来自正态总体的样本?

- 设总体分布是 $N(\mu,\sigma^2)$, 可以计算出 μ, σ^2 的最大似然估计分别是

$$\hat{\mu}=\bar{X}_n=403.5,\ \hat{\sigma}^2=\frac{1}{n}\sum_{j=1}^{n}(X_j-\hat{\mu})^2=83.12^2$$

- 数据的最小值是 $213>200$, 最大值是 $584<600$. 按照制作直方图的方法将将 $(200,600]$ 八等分. 计算出数据落入各段的频数为

$$(3,2,12,14,12,11,3,3)$$

- 用最大似然估计作为正态分布参数计算各个组的取值概率 $p_i, i=1,2,\ldots,m=8$。

- 各组频数和理论概率列表如下:

区间	频数f_j	理论概率p_j
(200, 250]	3	0.0252
(250, 300]	2	0.0741
(300, 350]	12	0.1534
(350, 400]	14	0.2233
(400, 450]	12	0.2289
(450, 500]	11	0.1651
(500, 550]	3	0.0838
(550, 600]	3	0.0300

- 这里第 1、2 组和第 7、8 组都频数太小，分别合并。

- 合并后的 6 个组的频数和理论概率列表如下:

区间	频数f_j	理论概率p_j
(200, 300]	5	0.0993
(300, 350]	12	0.1534
(350, 400]	14	0.2233
(400, 450]	12	0.2289
(450, 500]	11	0.1651
(500, 600]	6	0.1138

- 计算检验统计量

$$\xi=\sum_{j=1}^{6}\frac{(f_j-np_j)^2}{np_j}=0.1335$$

- 否定域为

$$\{\xi \geq \chi^2_{0.05}(6-1-2)\} = \{\xi \geq 7.815\}$$

- 检验统计量未落入否定域，在 0.05 水平下，不拒绝样本来自正态分布的假设。

第九章　线性回归分析

9.1　数据的相关性

- 二战初期, 德国对法国发动攻势后, 英国首相丘吉尔应法国的请求, 动用了十几个防空中队对德作战.
- 由于防空中队的飞机需要在欧洲大陆的机场进行维护, 使得空战中英国飞机损失严重.
- 这时, 法国总理请求英国继续增派十个中队的飞机, 丘吉尔决定同意这一要求.
- 英国内阁知道此事后, 请来统计学家利用线性回归模型对出动飞机与战损飞机的数据进行了统计分析, 发现如果飞机的补充率和损失率不变, 飞机数量的下降是非常快的:
- “以现在的损失率损失两周，英国在法国的飓风式战斗机就一架也不存在了”.
- 内阁希望丘吉尔收回他的决定.
- 最后，丘吉尔同意了内阁的要求，并在几天内撤回了在法国的飓风式战机，只留下了三个中队，为以后英国本土的保卫战保留了实力（读者 ·2006·4）.
- 线性回归方法是统计学中的常用方法, 是处理变量之间线性关系的重要方法.

数据的相关性—例

- 在实际问题中, 我们经常遇到有相关关系的变量. 比如讲身高与体重的关系时, 虽然身高不能确定体重, 但总的来讲, 身越高, 体越重.

- 在考虑某一个特定地区居民的身高和体重的关系时, 用 x 表示身高, 用 y 表示体重, 总体来讲, y 随着 x 增大一般也会增大. 这时我们称 y 和 x 有**相关关系**.

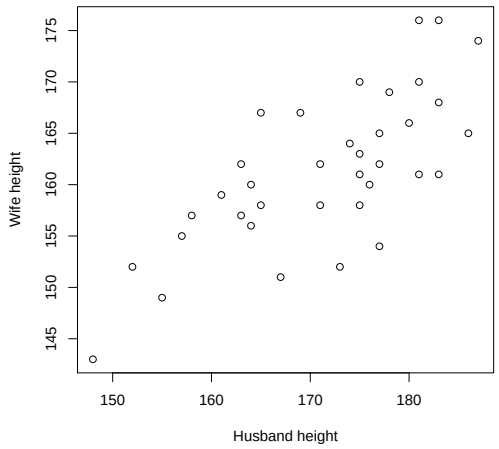
- **例 1.1** 一位社会学家对他所在城市进行了随机抽样调查, 得到了 36 对新婚夫妇的身高数据 (单位:cm).

- 数据对 (x_j, y_j) 中, x_j 是第 j 对夫妇中丈夫的身高, y_j 是妻子的身高. 我们称**数据对**

$$(x_j, y_j),\ j = 1, 2, \cdots, 36$$

为样本或观测数据. 这时, 样本是直角坐标系中的 36 个点, 将这 36 个点画在坐标系上得到观测数据的**散点图** (scatter plot). 横坐标是 x, 纵坐标是 y.

丈夫与妻子身高的散点图



9.1.1 样本相关系数

样本相关系数

- 无论是从抽样调查中得到的成对数据, 还是从科学试验, 工农业生产中得到的成对数据, 在统计学中也都称为观测数据或样本, 数据对的个数称为样本量.

- 样本量是 n 的成对观测数据是用

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n) \tag{1.1}$$

表示的.

- 这里, 对固定的 j, x_j 和 y_j 或来自相同的个体, 或是同一次试验的观测数据.
- 对 $i \neq j$, (x_i, y_i) 和 (x_j, y_j) 或来自不同的个体, 或是不同次试验的观测数据.
- 对于观测数据 (1.1), 我们用 $\{x_j\}$ 表示数据 $x_1, x_2, \cdots, x_n$, 用 $\{y_j\}$ 表示数据 $y_1, y_2, \cdots, y_n$.
- 用 $\bar{x}$ 和 $\bar{y}$ 分别表示 $\{x_j\}$ 和 $\{y_j\}$ 的样本均值.
- 用 s_x^2 表示 $\{x_j\}$ 的样本方差, 用 s_y^2 表示 $\{y_j\}$ 的样本方差.
- 对 $n > 2$, 有

$$\begin{array}{ll} \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, & s_x^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2, \\ \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j & s_y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2. \end{array}$$

- $s_x = \sqrt{s_x^2}$, $s_y = \sqrt{s_y^2}$ 分别是样本标准差.
- 再引入样本协方差

$$s_{xy} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y}). \tag{1.2}$$

- **定义 1.1**

  (1) 当 $s_x s_y \neq 0$, 我们称

  $$\hat{\rho}_{xy} = \frac{s_{xy}}{s_x s_y}$$

  为 $\{x_j\}$ 和 $\{y_j\}$ 的**样本相关系数**,

  (2) 当 $\hat{\rho}_{xy} > 0$, 我们称 $\{x_j\}$ 和 $\{y_j\}$ **正相关**,

  (3) 当 $\hat{\rho}_{xy} < 0$, 我们称 $\{x_j\}$ 和 $\{y_j\}$ **负相关**,

  (4) 当 $\hat{\rho}_{xy} = 0$, 我们称 $\{x_j\}$ 和 $\{y_j\}$ **不相关**.

- 大量的成对数据都显示了相关系数 $\hat{\rho}_{xy}$ 的以下性质 (参考习题 9.1):

  1 $\hat{\rho}_{xy}$ 总是在区间 $[-1, 1]$ 中取值,

  2 当 $|\hat{\rho}_{xy}| = 1$, 样本 $(x_j, y_j), j = 1, 2, \cdots, n$, 在同一条直线上,

  3 当 $\hat{\rho}_{xy}$ 接近于 1 时, x 增加, y 也倾向于增加, 这时数据 (1.1) 分散在一条上升的直线附近,

4 当 $\hat{\rho}_{xy}$ 接近于 -1 时, x 增加, y 倾向于减少, 这时数据 (1.1) 分散在一条减少的直线附近.

- 在实际问题中, 当 $|\hat{\rho}_{xy}| \geq 0.8$, 可以认为 $\{x_j\}$ 和 $\{y_j\}$ 高度相关;
- 当 $0.5 \leq |\hat{\rho}_{xy}| < 0.8$, 可以认为 $\{x_j\}$ 和 $\{y_j\}$ 中度相关;
- 当 $0.3 \leq |\hat{\rho}_{xy}| < 0.5$, 可以认为 $\{x_j\}$ 和 $\{y_j\}$ 低度相关;
- 当 $|\hat{\rho}_{xy}| < 0.3$, 可以认为 $\{x_j\}$ 和 $\{y_j\}$ 相关性极弱.
- 见演示。
- 容易计算出例 1.1 中丈夫和妻子身高的样本均值分别是 $\bar{x} = 171.3889$, $\bar{y} = 161.3333$, 样本标准差 $s_x = 9.9665$, $s_y = 7.4450$, 样本协方差 $s_{xy} = 53.8095$, 样本相关系数

$$\hat{\rho}_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{53.8095}{9.9665 \times 7.4450} \approx 0.7252.$$

  说明丈夫和妻子的身高是中度相关的.
- 粗略地讲, 丈夫的身高越高, 妻子的身高也会较高一点.
- 或说妻子身高越高, 丈夫的身高也会较高一点.

9.1.2 相关性检验

相关性检验

- 如果随机向量 (X_j, Y_j) 独立同分布, 并且和 (X, Y) 同分布, 就称 (X_j, Y_j) 及其观测值 (x_j, y_j), $j = 1, 2, \cdots, n$, 是来自总体 (X, Y) 的样本.
- 如果 (x_j, y_j), $j = 1, 2, \cdots, n$, 是来自总体 (X, Y) 的样本,

$$\rho_{xy} = \frac{\mathrm{E}[(X - \mathrm{E}X)(Y - \mathrm{E}Y)]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

  是 X, Y 的相关系数, 用强大数律可以证明样本相关系数

$$\hat{\rho}_{xy} = \frac{s_{xy}}{s_x s_y}$$

  是 ρ_{xy} 的强相合估计.

- 在实际问题中, 有时需要检验

$$H_0 : \rho_{xy} = 0 \quad \text{vs.} \quad H_1 : \rho_{xy} \neq 0. \tag{1.3}$$

- 如果否定了 H_0, 就认为 X, Y 是相关的.
- 当总体 (X, Y) 服从联合正态分布, H_0 成立时, 可以证明

$$T = \hat{\rho}_{xy}\sqrt{\frac{n-2}{1-\hat{\rho}_{xy}^2}} \sim t(n-2). \tag{1.4}$$

- 因为 H_0 成立时, $|\hat{\rho}_{xy}|$, 从而 $|T|$ 应当取值较小, 于是假设 (1.3) 的显著性水平为 α 的拒绝域是

$$W = \{|T| > t_{\alpha/2}(n-2)\}. \tag{1.5}$$

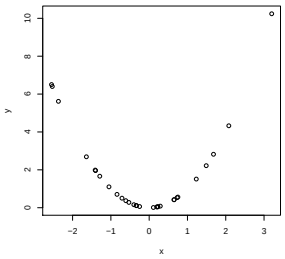
例 1.2

- 在例 1.1 中, $\hat{\rho}_{xy} = 0.7252$.
- 假设夫妻的身高总体 (X, Y) 服从联合正态分布.
- 经过计算得到

$$T = \hat{\rho}_{xy}\sqrt{\frac{n-2}{1-\hat{\rho}_{xy}^2}} = 6.1396.$$

- 查表得到 $t_{0.05/2}(36-2) = t_{0.025}(34) = 2.032 < T$.
- 检验的结果是显著的, 所以认为 X, Y 是相关的.
- 尽管拒绝域 (1.5) 是针对 X, Y 服从正态分布得到的, 但是对于非正态分布的情况, 当 n 较大, 人们也利用 (1.5) 作为假设 (1.3) 的拒绝域.
- 还应当指出, $\hat{\rho}_{xy} = 0$ 只是表示 X, Y 之间没有线性关系, 并不表示 X 和 Y 没有关系. 因为这时 X,Y 之间可能存在着非线性关系, 看下面的例子.

例 1.3



- 给定来自总体 X 的 30 个样本.
- 取 $y_j = x_j^2$, 则 (x_j, y_j) 是来自总体 $(X, Y) = (X, X^2)$ 的 30 个样本. 这些样本都在抛物线 $y = x^2$ 上, 因而 x, y 存在非线性函数关系. 见图形。
- 可以计算出

$$\hat{\rho}_{xy} = 0.0074.$$

- 这时,

$$T = \hat{\rho}_{xy}\sqrt{\frac{n-2}{1-\hat{\rho}_{xy}^2}} = 0.0392.$$

  查表得到 $t_{0.05/2}(30-2) = t_{0.025}(28) = 2.048 > T$.
- 所以不能否定 $H_0 : \rho_{xy} = 0$.
- 检验的 P 值是

$$P(|T_{28}| > 0.0392) = 2[1 - P(T_{28} < 0.0392)] \approx 0.97,$$

  所以应当承认 $H_0 : \rho_{xy} = 0$. 即 $\{x_j\}$, $\{y_j\}$ 不存在线性关系.

9.2 回归直线

回归直线

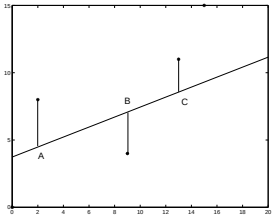
- 当 $\{x_j\}$ 和 $\{y_j\}$ 高度相关时, 我们已经知道数据

$$(x_1, y_1),\ (x_2, y_2),\ \cdots,\ (x_n, y_n)$$

  会分散在一条直线的附近.
- 我们将这条直线叫做回归直线, 下面就寻找这条直线.
- 在直角坐标系中, 两个点 (x_1, y_1), (x_2, y_2) 可以决定一条直线.

$$l: \qquad y = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1,\ \text{当 } x_2 \neq x_1.$$

  这时, 两个点都在直线上, 所以这两个点平均距离直线 l 最近.



- 给定三对数据,

$$(x_1, y_1),\ (x_2, y_2),\ (x_3, y_3),$$

当 x_1, x_2, x_3 不全相同, 我们也求一条直线 l, 使得以上三个点距离直线 l “平均最近”.

- 用

$$l: \qquad y = a + bx$$

表示要求的直线, 在平行于 y 轴的方向, 做以上三点到直线 l 的连线, 交点 A, B, C 的坐标分别是

$$A: (x_1, a + bx_1),\ B: (x_2, a + bx_2),\ C: (x_3, a + bx_3).$$

- 三对观测数据和交点 A, B, C 的距离分别是

$$|y_1 - (a + bx_1)|,\ |y_2 - (a + bx_2)|,\ |y_3 - (a + bx_3)|.$$

- 我们用这三个距离的平方和

$$(y_1 - a - bx_1)^2 + (y_2 - a - bx_2)^2 + (y_3 - a - bx_3)^2.$$

衡量这三对观测数据远离直线 l 的程度.

- 如果 a, b 使得

$$Q(a, b) = (y_1 - a - bx_1)^2 + (y_2 - a - bx_2)^2 + (y_3 - a - bx_3)^2$$

达到最小就称直线 $l:\ y = a + bx$ 是数据的回归直线.

- 一般地要为样本量是 n 的观测数据,

$$(x_1, y_1),\ (x_2, y_2),\ \cdots,\ (x_n, y_n),\ \text{其中的 } x_j \text{ 不全相同,}$$

建立回归直线 $l:\ y = a + bx$, 使之与观测数据平均最近时, 也采用相同的方法.

- 沿平行于 y 轴的方向, 点 (x_j, y_j) 到它与 l 的交点的距离是

$$|y_j - (a + bx_j)|,\ j = 1, 2, \cdots, n.$$

- 我们用这些距离的平方和

$$\begin{aligned} Q(a,b) =& (y_1 - a - bx_1)^2 + (y_2 - a - bx_2)^2 + \cdots \\ &+ (y_n - a - bx_n)^2 \end{aligned} \tag{2.1}$$

衡量观测数据远离直线 l 的程度.

- 如果常数 a, b 使得 $Q(a,b)$ 达到最小, 就称直线

$$l: \qquad y = a + bx$$

是 $\{x_j\}$ 与 $\{y_j\}$ 的**回归直线** (regression line).

- 得到了回归直线后, 只要 $\{x_j\}$ 与 $\{y_j\}$ 相关性较强, 对于新的 x, 就可以用回归直线上的点 $\hat{y} = a + bx$ 作为 y 的预测值.

- 事实证明: $|\hat{\rho}_{xy}|$ 越接近于 1, 预测就越准确; x 越接近 $\bar{x}$, 预测也越好.

最小二乘解

- **定理 2.1** 如果 $\{x_j\}$ 不全相同, 则 $Q(a,b)$ 的最小值点是

$$\begin{aligned} \hat{b} =& \frac{s_{xy}}{s_x^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \\ \hat{a} =& \bar{y} - \hat{b}\bar{x}. \end{aligned} \tag{2.2}$$

- 称 $\hat{a}, \hat{b}$ 是回归直线系数 a, b 的**最小二乘估计**。

- **证明** 将 $Q(a,b)$ 右端的每一项对 a, b 进行二项展开, 就知道 $Q(a,b)$ 是 a, b 的二元二次多项式, 二次项系数大于零.

- $Q(a,b)$ 是开口向上的椭圆抛物面, 最小值唯一存在, 并且可以令两个一阶偏导数为零得到最小值点.

- 由

$$\begin{aligned}\frac{\partial Q}{\partial a} =& -2\sum_{j=1}^{n}(y_j - a - bx_j) = 0,\\ \frac{\partial Q}{\partial b} =& -2\sum_{j=1}^{n}(y_j - a - bx_j)x_j = 0,\end{aligned} \tag{2.3}$$

- 得到方程组

$$\begin{aligned}\bar{y} - a - b\bar{x} =& 0;\\ \sum_{j=1}^{n} y_j x_j - an\bar{x} - b\sum_{j=1}^{n} x_j^2 =& 0,\end{aligned}$$

- 把前式中 a 代入后式，并利用

$$\begin{aligned}\sum_{j=1}^{n} y_j x_j - n\bar{x}\,\bar{y} =& \sum_{j=1}^{n}(y_j - \bar{y})(x_j - \bar{x}),\\ \sum_{j=1}^{n} x_j^2 - n\bar{x}\,\bar{x} =& \sum_{j=1}^{n}(x_j - \bar{x})^2\end{aligned}$$

- 得

$$\begin{aligned}\sum x_j y_j - n(\bar{y} - b\bar{x})\bar{x} - b\sum x_j^2 =& 0\\ \sum x_j y_j - n\bar{x}\bar{y} - b(\sum x_j^2 - n\bar{x}^2) =& 0\end{aligned}$$

- 解出

$$\begin{aligned}\hat{b} =& \frac{\sum_{j=1}^{n} y_j x_j - n\bar{x}\,\bar{y}}{\sum_{j=1}^{n} x_j^2 - n\bar{x}^2} = \frac{s_{xy}}{s_x^2},\\ \hat{a} =& \bar{y} - \hat{b}\bar{x}.\end{aligned}$$

- 于是直线

$$l: \quad \hat{y} = \hat{a} + \hat{b}x$$

为数据 $(x_j, y_j), j = 1, 2, \cdots, n$ 的回归直线.

- 给了任何另外的 x, 当 $|\hat{\rho}_{xy}|$ 接近于 1 时, 我们可以用回归直线 l 上的点

$$\hat{y} = \hat{a} + \hat{b}x \tag{2.4}$$

对和 x 相应的 y 作出预测.

- 我们又称 (2.4) 为**经验公式**.
- 从 (2.2) 的第二式知道, $(\bar{x}, \bar{y})$ 总在回归直线上.
- 因为 $\hat{a}, \hat{b}$ 是极小化 a, b 的二次函数 $Q(a, b)$ 得到的, 所以又称它们是 a, b 的**最小二乘估计** (least square estimator).

例 2.1

- 容易计算例 1.1 中的 $\bar{x} = 171.39, \bar{y} = 161.33$,

$$s_x^2 = 99.3302,\ s_{xy} = 53.8095,$$

- 于是得到

$$\hat{b} = \frac{53.8095}{99.3302} = 0.5417,\ \hat{a} = \bar{y} - \hat{b}\bar{x} = 68.4920.$$

- 回归直线是

$$\hat{y} = 68.492 + 0.5417x.$$

9.3 一元线性回归

一元线性回归–模型

- 对数据 (x_j, y_j) 建立了回归直线 l 后, 我们用回归直线 l 上的

$$\hat{y}_j = \hat{a} + \hat{b}x_j$$

作为 y_j 的预测值, **预测误差**是

$$\hat{\varepsilon}_j = y_j - \hat{y}_j = y_j - \hat{a} - \hat{b}x_j.$$

- 我们也称 $\hat{\varepsilon}_j$ 为残差, 称残差的平方和

$$Q = \sum_{j=1}^{n} \hat{\varepsilon}_j^2 = \sum_{j=1}^{n} (y_j - \hat{a} - \hat{b}x_j)^2 = Q(\hat{a}, \hat{b}) \tag{3.1}$$

为**残差平方和**.

- Q 较小时, l 代表了 x, y 之间的线性关系:

$$y_j = \hat{a} + \hat{b}x_j + \hat{\varepsilon}_j,\ j = 1, 2, \cdots, n.$$

- 为了统计分析的方便, 我们认为成对数据 (x_j, y_j) 满足模型

$$Y_j = a + bx_j + \varepsilon_j,\ j = 1, 2, \cdots, n. \tag{3.2}$$

- 其中的 a, b 是未知常数, $\{\varepsilon_j\}$ 是独立同分布的随机变量, 服从正态分布 $N(0, \sigma^2)$.
- 这里 σ^2 是未知正数, 代表了随机误差的强弱. σ^2 越大, 说明随机误差越强.
- 模型 (3.2) 称为**一元线性回归模型** (linear regression model).
- 其中 a, b 分别是直线 $y = a + bx$ 的截距和斜率, 称为**回归参数**.
- 在模型 (3.2) 中, 我们称 x_j 是设计变量或输入变量, 它表示得到 Y_j 时的输入条件.
- 我们将 x_j 看作常量, 不做随机变量处理.
- Y_j 是观测变量, 它是输入条件 x_j 后得到的观测结果. 我们称 (x_j, y_j) 是来自一元线性回归模型的样本.
- 给定来自一元线性回归模型 (3.2) 的样本 (x_j, y_j), 我们的问题是需要估计出回归参数 a, b 和 σ^2.

9.3.1 最大似然估计和最小二乘估计

回归参数的最大似然估计

- 设 (x_j, Y_j) 满足一元线性回归模型 (3.2), 则 Y_j, $j = 1, 2, \cdots, n$ 相互独立, 都服从正态分布.
- 利用

$$\mathrm{E}Y_j = a + bx_j,\ \mathrm{Var}(Y_j) = \mathrm{Var}(\varepsilon_j) = \sigma^2,$$

  知道 $Y_j \sim N(a + bx_j, \sigma^2)$.
- 于是得到基于 $Y_1, Y_2, \cdots, Y_n$ 的似然函数

$$\begin{aligned} L(a, b, \sigma^2) =& \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(y_j - a - bx_j)^2 \Big) \\ =& \Big(\frac{1}{\sqrt{2\pi\sigma^2}} \Big)^n \exp\Big(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} (y_j - a - bx_j)^2 \Big) \\ =& \Big(\frac{1}{\sqrt{2\pi\sigma^2}} \Big)^n \exp\Big(-\frac{1}{2\sigma^2} Q(a, b) \Big). \end{aligned} \tag{3.3}$$

  其中的 $Q(a, b)$ 由 (2.1) 定义.

- 对数似然函数是

$$l(a,b,\sigma^2) = -\frac{1}{2\sigma^2}Q(a,b) - \frac{n}{2}\ln\sigma^2 - n\ln\sqrt{2\pi}.$$

- 解方程组

$$\begin{cases} \frac{\partial l}{\partial a} = -\frac{1}{2\sigma^2}\frac{\partial Q}{\partial a} = 0, \\ \frac{\partial l}{\partial b} = -\frac{1}{2\sigma^2}\frac{\partial Q}{\partial b} = 0, \\ \frac{\partial l}{\partial \sigma^2} = \frac{1}{2\sigma^4}Q(a,b) - \frac{n}{2\sigma^2} = 0, \end{cases}$$

  可以得到 a, b, σ^2 的最大似然估计.

- 注意前两个方程和 (2.3) 是等价的, 所以当 $s_x^2 \neq 0$, 从前两个方程得到 a, b 的最大似然估计

$$\hat{b} = \frac{s_{xy}}{s_x^2},\ \hat{a} = \bar{y} - \hat{b}\bar{x}. \tag{3.4}$$

- 将 $\hat{a}, \hat{b}$ 代入第三个方程, 得到 σ^2 的最大似然估计是 $\frac{1}{n}Q(\hat{a},\hat{b})$.

- 因为数学上可以证明

$$\mathrm{E}Q(\hat{a},\hat{b}) = (n-2)\sigma^2,$$

  所以 σ^2 的最大似然估计不是 σ^2 无偏估计, 为了使用无偏估计, 我们以后用

$$\hat{\sigma}^2 = \frac{1}{n-2}Q(\hat{a},\hat{b}) \tag{3.5}$$

  作为 σ^2 的估计. 这时有 $\mathrm{E}\hat{\sigma}^2 = \sigma^2$.

- 容易看出, (a, b) 的最小二乘估计和最大似然估计是相同的.

- 引入记号:

$$\begin{aligned} l_{yy} &\triangleq \sum_{i=1}^{n}(y_i - \bar{y})^2 \\ l_{xx} &\triangleq \sum_{i=1}^{n}(x_i - \bar{x})^2 \\ l_{xy} &\triangleq \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ Q &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \end{aligned}$$

- 则

$$\hat{b} = \frac{l_{xy}}{l_{xx}}, \qquad \hat{\sigma}^2 = \frac{1}{n-2}Q$$

回归参数估计的性质

- 以后总设 $s_x^2 \neq 0$. 下面是这些估计量的基本性质.

- **定理 3.1** 设 (x_j, Y_j), $j = 1, 2, \cdots, n$, 满足一元线性回归模型 (3.2), $\hat{a}, \hat{b}, \hat{\sigma}^2$ 由 (3.4) 和 (3.5) 定义, $n > 2$ 时, 有

  (1)
  $$\hat{b} \sim N\Big(b, \frac{\sigma^2}{l_{xx}}\Big),$$

  (2)
  $$\hat{a} \sim N\Big(a,\ [\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}}]\sigma^2\Big),$$

  (3)
  $$\frac{n-2}{\sigma^2}\hat{\sigma}^2 \sim \chi^2(n-2),$$

  (4)
  $$\bar{Y}, \hat{b}, \hat{\sigma}^2 \text{ 相互独立.}$$

定理证明

- (1) 对任何常数 c, 有
$$\sum_{j=1}^{n}(x_j - \bar{x})c = (n\bar{x} - n\bar{x})c = 0. \tag{3.6}$$

- 利用 (3.6) 得到
$$\begin{aligned}\hat{b} =& \frac{1}{s_x^2}\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})(Y_j - \bar{Y}) \\ =& \frac{1}{s_x^2}\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})Y_j. \end{aligned} \tag{3.7}$$

- 由于 $\hat{b}$ 已经是相互独立的正态随机变量 Y_j 的线性组合, 所以服从正态分布.

- 只需要再计算它的数学期望和方差.

- 利用 $Y_j \sim N(a+bx_j, \sigma^2)$, (3.6) 和 (3.7), 得到

$$\begin{aligned}\mathrm{E}\hat{b} &= \frac{1}{s_x^2}\frac{1}{n-1}\sum_{j=1}^{n}(x_j-\bar{x})\mathrm{E}Y_j \\ &= \frac{1}{s_x^2}\frac{1}{n-1}\sum_{j=1}^{n}(x_j-\bar{x})(a+bx_j) \\ &= \frac{1}{s_x^2}\frac{b}{n-1}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x}) \\ &= \frac{b}{s_x^2}s_x^2 = b.\end{aligned}$$

- 因为 $Y_1, Y_2, \cdots, Y_n$ 相互独立, 有共同的方差 σ^2, 利用 (3.7) 得到

$$\begin{aligned}\mathrm{Var}(\hat{b}) =& \frac{1}{s_x^4}\frac{1}{(n-1)^2}\sum_{j=1}^{n}(x_j-\bar{x})^2\mathrm{Var}(Y_j) \\ =& \frac{1}{s_x^4}\frac{1}{(n-1)}s_x^2\sigma^2 = \frac{\sigma^2}{(n-1)s_x^2} = \frac{\sigma^2}{l_{xx}}.\end{aligned}$$

- (2) 设 $\bar{\varepsilon}_n = \sum_{j=1}^{n}\varepsilon_j/n$, 在 (3.2):

$$Y_j = a + bx_j + \varepsilon_j,\ j = 1, 2, \cdots, n. \tag{3.2}$$

  两边求样本平均得到

-

$$\bar{Y} = a + b\bar{x} + \bar{\varepsilon}_n,\ \ \mathrm{E}\bar{Y} = a + b\bar{x}. \tag{3.8}$$

- 因为 $\bar{Y}$ 和 $\hat{b}$ 都是 $Y_1, Y_2, \cdots, Y_n$ 的线性组合, 所以 $\hat{a} = \bar{Y} - \hat{b}\bar{x}$ 也是 $Y_1, Y_2, \cdots, Y_n$ 的线性组合, 从而也服从正态分布.
- 再计算它的均值和方差如下.
- 利用 (3.8) 得到

$$\begin{aligned}\mathrm{E}\hat{a} =& \mathrm{E}(\bar{Y} - \hat{b}\bar{x}) = \mathrm{E}\bar{Y} - \mathrm{E}\hat{b}\bar{x} \\ =& a + b\bar{x} - b\bar{x} = a.\end{aligned}$$

- 最后再利用

$$\hat{a} = \bar{Y} - \hat{b}\bar{x}$$

  得到

$$\begin{aligned}\mathrm{Var}(\hat{a}) =& \mathrm{Var}(\bar{Y}) + \mathrm{Var}(\hat{b}\bar{x}) - 2\bar{x}\mathrm{Cov}(\bar{Y}, \hat{b}) \\ =& \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{(n-1)s_x^2} - 2\bar{x}\mathrm{Cov}(\bar{Y}, \hat{b}).\end{aligned}$$

- 由于从 (3.7) 得到

$$\begin{aligned}\mathrm{Cov}(\bar{Y},\hat{b}) =&\mathrm{Cov}\Big(\frac{1}{n}\sum_{j=1}^{n}Y_j,\ \frac{1}{(n-1)s_x^2}\sum_{j=1}^{n}(x_j-\bar{x})Y_j\Big)\\ =&\frac{1}{n(n-1)s_x^2}\mathrm{Cov}\Big(\sum_{j=1}^{n}Y_j,\ \sum_{j=1}^{n}(x_j-\bar{x})Y_j\Big)\\ =&\frac{1}{n(n-1)s_x^2}\sum_{j=1}^{n}(x_j-\bar{x})\sigma^2\\ =&0,\end{aligned}$$

- 这说明 $\bar{Y}$ 和 $\hat{b}$ 独立。

- 于是有

$$\mathrm{Var}(\hat{a}) = \Big(\frac{1}{n}+\frac{\bar{x}^2}{(n-1)s_x^2}\Big)\sigma^2.$$

- 性质 (3)、(4) 证明略去。

9.3.2 平方和分解公式

平方和分解公式

- 最小二乘回归直线：$l:\quad \hat{y}=\hat{a}+\hat{b}x$

- 总平方和为因变量 y 的变动情况，代表了 y 包含的信息多少：

$$l_{yy}\triangleq\sum_{i=1}^{n}(y_i-\bar{y})^2$$

- 残差平方和代表用直线解释 y 与 x 间关系的接近程度:

$$Q\triangleq\sum_{i=1}^{n}(y_i-\hat{y})^2$$

- 回归平方和是由 x 的值所确定的 y 的变化情况:

$$l_{\hat{y}\hat{y}}\triangleq\sum_{j=1}^{n}(\hat{y}_i-\bar{y})^2$$

- 注意

$$\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i=\frac{1}{n}\sum_{i=1}^{n}(\hat{a}+\hat{b}x_i)=\hat{a}+\hat{b}\bar{x}=\bar{y}$$

- 另外

$$l_{\hat{y}\hat{y}} = \sum_{i=1}^{n}[(\hat{a}+\hat{b}x_i)-(\hat{a}+\hat{b}\bar{x})]^2 = \sum_{i=1}^{n}[\hat{b}(x_i-\bar{x})]^2 = \hat{b}^2 l_{xx}$$

其中

$$l_{xx} \triangleq \sum_{i=1}^{n}(x_i-\bar{x})^2$$

- 总平方和可以分解为回归平方和与残差平方和：

$$l_{yy} = l_{\hat{y}\hat{y}} + Q = \hat{b}^2 l_{xx} + Q \tag{3.12}$$

- 证明只要看

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

并证明交叉项为零:

$$\begin{aligned}&\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) = \sum_{i=1}^{n}[(y_i-\bar{y})-(\hat{y}_i-\bar{y})](\hat{y}_i-\bar{y})\\ =&\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{y}) - \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2\\ =&\hat{b}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x}) - \hat{b}^2\sum_{i=1}^{n}(x_i-\bar{x})^2 = 0\end{aligned}$$

9.3.3 斜率 b 的检验

斜率 b 的检验

- 当斜率 $b=0$ 时，回归直线退化为 $l:\ y=a$，这时 x 对 y 没有影响，线性回归没有意义。
- 只有 $b\neq 0$ 的回归结果才是有意义的。
- 检验:

$$H_0:\ b=0 \qquad \text{vs.} \qquad H_a:\ b\neq 0$$

- 直观构造否定域: 当 $|\hat{b}|$ 很大时拒绝 H_0。
- 需要知道 $\hat{b}$ 在 H_0 下的分布才能给出“很大”的临界值。
- 定理 3.1 说明 $\hat{b}\sim \mathrm{N}(b,\sigma^2/l_{xx})$; $(n-2)\hat{\sigma}^2/\sigma^2\sim\chi^2(n-2)$ 且与 $\hat{b}$ 独立，说明在 H_0 下

$$T = \frac{\hat{b}}{\hat{\sigma}/\sqrt{l_{xx}}} \sim \mathrm{t}(n-2)$$

- 于是水平 α 的否定域为

$$W = \{|T| > t_{\alpha/2}(n-2)\}$$

回归系数检验的例子

- 对例 1.1 的数据，检验 b 是否等于零。
- 计算得 $\hat{b} = 0.5417$, $\hat{\sigma}^2 = 27.0538$, $l_{xx} = 3476.6$, $n = 36$, $t_{0.025}(34) = 2.032$, $T = 6.14$ 落入否定域，应拒绝 $H_0 : b = 0$。

9.3.4 预测的置信区间

预测的置信区间

- 回归直线 $l : \hat{y} = \hat{a} + \hat{b}x$。
- 对于新的输入条件 x_0, 预测

$$\hat{y}_0 = \hat{a} + \hat{b}x_0$$

- 未知的真实值为

$$Y_0 = a + bx_0 + \varepsilon_0$$

- 要得到 Y_0 的置信区间 (预测区间), 必须知道 $Y_0 - \hat{y}_0$ 的概率分布.
- 引入

$$\eta_0 = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}}}.$$

- **定理 3.2** 如果 (x_j, y_j), $j = 1, 2, \cdots, n$, 是来自一元线性回归模型 (3.2) 的数据, 则

$$\frac{Y_0 - \hat{y}_0}{\eta_0} \sim t(n-2). \tag{3.21}$$

- 利用定理 3.2 和

$$\begin{aligned} & P\big(\hat{y}_0 - t_{\alpha/2}(n-2)\eta_0 \le Y_0 \le \hat{y}_0 + t_{\alpha/2}(n-2)\eta_0\big) \\ =& P\Big(\frac{|Y_0 - \hat{y}_0|}{\eta_0} \le t_{\alpha/2}(n-2)\Big) \\ =& 1 - \alpha, \end{aligned}$$

- 得到 Y_0 的置信度为 $1-\alpha$ 的置信区间

$$[\,\hat{y}_0 - t_{\alpha/2}(n-2)\eta_0,\ \hat{y}_0 + t_{\alpha/2}(n-2)\eta_0\,]. \tag{3.22}$$

- 置信区间的长度是

$$\begin{aligned} L =& 2t_{\alpha/2}(n-2)\eta_0 \\ =& 2t_{\alpha/2}(n-2)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}}. \end{aligned}$$

- 在相同的置信度下, 置信区间的长度越小越好.

- 相同的置信度下, n 越大, L 越小;

- l_{xx} 越大, L 越小;

- x_0 离 $\bar{x}$ 越近,L 越小;

- $\hat{\sigma}$ 越小, L 越小.

例 3.3

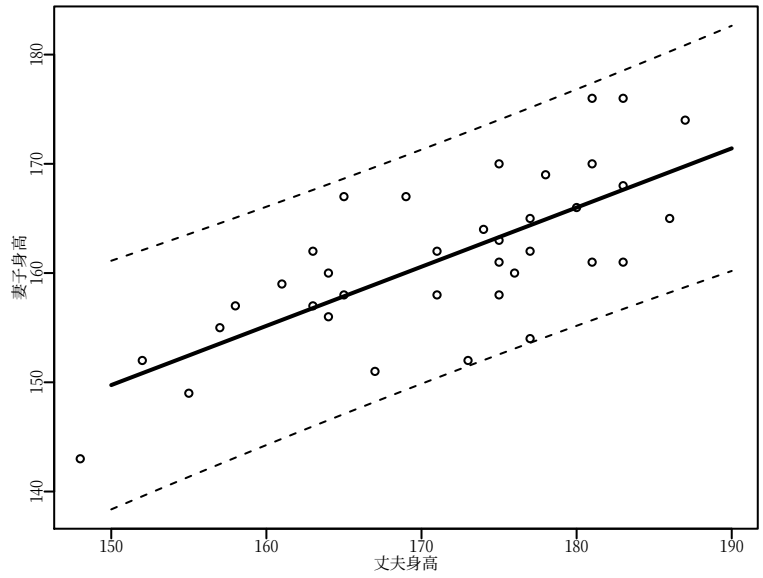
- 在例 1.1 中，令 $x_0 \in [150, 190]$, 计算 $\hat{y}_0$ 及 Y_0 的水平为 0.95 的预测区间。

- 解答：这里 $n = 36$, $t_{0.05/2}(34) = 2.032$, $\hat{y}_0 = 68.492 + 0.5417x_0$, $\hat{\sigma}_0 = \sqrt{27.0538} = 5.2013$,

$$\begin{aligned} \eta_0 =& t_{0.05/2}(34)\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}} \\ =& 10.5690\sqrt{1.0278+\frac{(x_0-171.39)^2}{3476.6}} \end{aligned}$$

  令 $\hat{y}_0^+ = \hat{y}_0 + \eta_0$, $\hat{y}_0^- = \hat{y}_0 - \eta_0$。

- 绘制 (x_j, y_j) 的散点图，$(x_0, \hat{y}_0)$ 的变化曲线，$(x_0, \hat{y}_0^-)$ 与 $(x_0, \hat{y}_0^+)$ 的变化曲线如下。



9.4 多元线性回归

多元线性回归

- 一元线性回归的模型可以推广到有多个自变量（设计变量）的情形。
- 模型

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p + \varepsilon. \tag{4.1}$$

- 其中 $b_0, b_1, \ldots, b_p$ 是未知常数，称为**回归系数**。ε 是随机变量，$E\varepsilon = 0$, $\mathrm{Var}(\varepsilon) = \sigma^2$, σ^2 是未知参数。
- 称 (4.1) 是多元线性回归模型。

- 如果有 n 组观测值满足

$$Y_j = b_0 + b_1 x_{j1} + \cdots + b_p x_{jp} + \varepsilon_j, \quad j = 1, 2, \ldots, n \tag{4.2}$$

且 $\varepsilon_j, j = 1, 2, \ldots, n$ 独立同 $\mathrm{N}(0, \sigma^2)$ 分布，则称

$$(Y_j; x_{j1}, x_{j2}, \ldots, x_{jp}), j = 1, 2, \ldots, n \tag{4.3}$$

为来自 p 元线性回归模型 (4.1) 的样本。

- 这时 $Y_1, Y_2, \ldots, Y_n$ 相互独立，且

$$Y_j \sim \mathrm{N}(b_0 + b_1 x_{j1} + \cdots + b_p x_{jp},\ \sigma^2). \tag{4.4}$$

9.4.1 最小二乘估计

最小二乘估计

- 令

$$Q(b_0, b_1, \ldots, b_p) = \sum_{j=1}^{n} \left[y_j - (b_0 + b_1 x_{j1} + \cdots + b_p x_{jp}) \right]^2 \tag{4.5}$$

- 如果 $(\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_p)$ 是 $Q(b_0, b_1, \ldots, b_p)$ 的最小值点，称 $(\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_p)$ 是回归系数 $(b_0, b_1, \ldots, b_p)$ 的**最小二乘估计**。

矩阵表示

- 将 (4.2) 的 n 个方程写成矩阵形式：

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},\ \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix},$$

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix},\ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{\varepsilon}.$$

- 这时

$$Q(b_0, b_1, \ldots, b_p) = Q(\boldsymbol{b}) = \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}\|^2. \tag{4.7}$$

- 用求偏导数的方法或者用线性代数投影的方法，可以证明最小二乘估计存在，且当 $\boldsymbol{X}^T\boldsymbol{X}$ 可逆时，估计为

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}. \tag{4.9}$$

预测

- 令

$$\begin{aligned}\hat{y}_j =&\hat{b}_0 + \hat{b}_1 x_{j1} + \cdots + \hat{b}_p x_{jp}, \\ \hat{\varepsilon}_j =&y_j - \hat{y}_j,\ j = 1, 2, \ldots, n\end{aligned} \tag{4.10}$$

  称 $\hat{y}_j$ 为**预测值**或**拟合值**，称 $\hat{\varepsilon}_j$ 为**残差**。

- 称

$$Q = \sum_{j=1}^{n} \hat{\varepsilon}_j^2 = Q(\hat{\boldsymbol{b}})$$

  为**残差平方和**。

- σ^2 的无偏估计为

$$\hat{\sigma}^2 = \frac{1}{n-p-1} Q. \tag{4.11}$$

- 回归的经验公式为

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \cdots + \hat{b}_p x_p.$$

参数估计的性质

- **定理 4.1** 设观测数据 (4.3) 是来自 p 元线性回归模型 (4.1) 的样本，则

$$\begin{aligned}&(1)\ E\hat{\sigma}^2 = \sigma^2, \\ &(2)\ \hat{\boldsymbol{b}} \sim \mathrm{N}(\boldsymbol{b}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}), \\ &(3)\ \frac{(n-p-1)}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p-1}, \\ &(4)\ \hat{\boldsymbol{b}}\text{与}\hat{\sigma}^2\text{独立}.\end{aligned}$$

9.4.2 回归显著性检验

回归显著性检验

- 最小二乘估计总存在，但是解出的经验公式是否有意义？

- 检验

$$H_0:\ b_1 = \cdots = b_p = 0$$

若 H_0 成立，模型退化成 $Y = b_0 + \varepsilon$，不出现自变量。

- 只有否定了 H_0，回归结果才有意义。
- 这个检验也称为方差分析检验，所有斜率项等于零的检验。

平方和分解

- 令 $l_{yy} = \sum_{j=1}^{n}(y_j - \bar{y})^2$, $\hat{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{b}$, $Q = \sum_{j=1}^{n}(y_j - \hat{y}_j)^2 = \|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|^2$, $l_{\hat{y}\hat{y}} = \sum_{j=1}^{n}(\hat{y}_j - \bar{y})^2$, 仍有平方和分解公式

$$l_{yy} = l_{\hat{y}\hat{y}} + Q.$$

- 为使得回归模型有意义，分解中回归平方和 $l_{\hat{y}\hat{y}}$ 越大越好，残差平方和 Q 越小越好。

回归显著性检验方法

- 定义统计量

$$F = \frac{l_{\hat{y}\hat{y}}/p}{Q/(n-p-1)}$$

在 H_0 成立时 $F \sim F(p, n-p-1)$。

- 取 $F(p, n-p-1)$ 上侧 α 分位数 $F_\alpha(p, n-p-1)$, 当 $F > F_\alpha(p, n-p-1)$ 拒绝 H_0。

9.4.3 单个系数的显著性检验

单个系数的显著性检验

- 对某个自变量 x_k，如果 $b_k = 0$，则 x_k 不出现在模型中。
- 考虑

$$H_0:\ b_k = 0$$

的检验。

- 由定理 4.1, 设 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ 的第 $(k+1, k+1)$ 元素为 c_{kk}，则 $\hat{\beta}_k \sim \mathrm{N}(b_k, c_{kk}\sigma^2)$，$\frac{(n-p-1)}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-p-1}$, 且 $\hat{\beta}_k$ 与 $\hat{\sigma}^2$ 独立。

- 定义统计量

$$T_k = \frac{\hat{b}_k}{\hat{\sigma}\sqrt{c_{kk}}},$$

则 H_0 成立时 $T_k \sim t(n-p-1)$。

- 取 $t(n-p-1)$ 分布的双侧 α 分位数 $t_{\alpha/2}(n-p-1)$, 当 $|T| \geq t_{\alpha/2}(n-p-1)$ 拒绝 H_0。

- 当 $|T| < t_{\alpha/2}(n-p-1)$ 不拒绝 H_0，这时可以从模型中将自变量 x_k 剔除，用剩余的自变量去估计模型。这叫做自变量选择。

9.4.4 残差诊断

标准化残差

- 残差 $\hat{\varepsilon}_j$ 作为误差 ε 的估计，其分布与 $\mathrm{N}(\sigma^2)$ 有较大差别。

- 实际上，

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}} =& \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{b}} \\ =& \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T) = \boldsymbol{P}\boldsymbol{Y}, \\ \hat{\boldsymbol{\varepsilon}} \sim& \mathrm{N}(0, \sigma^2\boldsymbol{P}).\end{aligned}$$

- 记 $n \times n$ 矩阵 $\boldsymbol{P}$ 的第 (j,j) 元素为 p_{jj}，则 $\hat{\varepsilon}_j \sim \mathrm{N}(0, \sigma^2 p_{jj})$，方差不相同。

- 作如下标准化：

$$\hat{e}_j = \frac{\hat{\varepsilon}_j}{\hat{\sigma}\sqrt{p_{jj}}}$$

则当 n 较大时 $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n$ 近似服从独立的标准正态分布。大部分标准化残差应该位于正负 2 之间。

- 可以画残差的各种图形，以检查模型设定是否合理，称为残差诊断。