经济与管理学的数学基础

概率论与数理统计部分讲义

复旦大学管理学院统计学系 徐勤丰

2009

概率论与数理统计部分教学目标

系统回顾以微积分、线性代数为基础的概率论与数理统计的基础知识。

教材与参考书目

Casella, G. and Berger, R. L., *Statistical Inference* (2nd edition), Wadsworth, 2001. （影印版，机械工业出版社，2002）

茆诗松等，概率论与数理统计教程，高等教育出版社，2004。

茆诗松等，高等数理统计，高等教育出版社，1998。

课程安排：

概率论部分四次，数理统计部分四次。

1．第一章 概率；第二章 随机变量

2．第三章 随机向量

3．第四章 数字特征

4．第五章 特征函数；第六章 概率极限定理

5．第七章 样本与统计量

6．第八章 参数估计（点估计）

7．第八章 参数估计（区间估计）；第九章 假设检验 （基本概念）

8．第九章 假设检验

考核方式：

平时作业 （30%） 考试（70%）

第一章　概 率

随机现象

- 特点：
  a) 可能结果不止一个；
  b) 究竟出现哪个结果，事先无法预知，视机会而定。

- 有规律可循，其规律称为“统计规律”。不同随机现象的统计规律有共性。

概率论

- 起源于赌博，17 世纪 Pascal 、Fermat、Huygens 等人的研究。
- 是研究随机现象中的统计规律的一门学科。提供了一套用于对总体、随机现象建模的工具。是统计学的基础。

§1.1 概率的公理化定义

1. 基本概念

概率论建立于集合论的基础之上。

随机试验(Random Experiment)

在一定条件下（自然条件或实验条件），对某种随机现象进行的一次观测，称为一次随机试验。基本特点是：所有可能的观测结果事先已知，但一次试验中究竟观测到哪一个具体结果，事先无法预知，视机会而定。

基本概念	概率论的含义	集合论的语言
样本点	一项随机试验可能出现的基本结果	元素，常用 s, w 等符号表示
样本空间	所有基本结果的集合	全集，我们用 S 表示，文献中也常用 Ω 表示
随机事件	对试验结果的判断或陈述	子集，常用 A，B，C 等大写字母表示
基本事件	仅涉及单个基本结果的陈述	由一个元素构成的子集

事件的发生、不发生:

设试验的实际结果为 s，若 $s \in A$，则称事件 A 发生；否则称 A 不发生。

S 为必然事件，$\varnothing$ 为不可能事件。

例 1.1

1) 掷一枚硬币一次。

可能的样本点为：s_1：得正面，s_2：得反面。样本空间为 $S=\{s_1, s_2\}$。

2) 从一批产品任取三个，检查是否合格。

若关心每件产品是否合格，样本空间为 $S=\{(GGG),(GGB),\mathbf{L},(BBB)\}$，包含 8 个样本点；

若只关心不合格产品数，则样本空间可取为 $S=\{0,1,2,3\}$。

3) 观察某天早上 7～8 点间到达复旦车站的乘客数。

S 可取为非负整数集。

4) 观察某只股票明天的收盘价。

$S=\{s:\ a(1-10\%)\le s\le a(1+10\%)\}$，其中 a 为今天的收盘价，已知。

注：

- 样本空间的选取应根据研究目的而定，也往往需要考虑数学处理的方便程度。
- 不同的随机现象可能用相同的样本空间去描述。

事件的关系与运算

- 包含：$A\subset B$，含义为“A 发生则 B 必发生”。
- 相等：$A=B \Longleftrightarrow A\subset B$ 且 $B\subset A$。
- 交：A 与 B 同时发生，记作 $A\mathbf{I}B=AB=\{s: s\in A\text{且}\ s\in B\}$。
- 并：A 或 B 发生，记作 $A\mathbf{U}B=\{s: s\in A\ \text{或}\ s\in B\}$。

  若 $AB=\varnothing$，称 A 与 B 不相容（互斥），此时 $A\mathbf{U}B$ 也记作 $A+B$。
- 余：A 的对立事件（A 不发生）称为 A 的余，记作 $A^c=\{s:\ s\overline{\in}A\}$。

  显然，$AA^c=\varnothing,\quad A+A^c=S$。
- 差：A 发生但 B 不发生，记作 $A-B=AB^c$。

  $A^c=S-A$。

运算规则

1) 交换律：$A \cup B = B \cup A,\ AB = BA$。

2) 结合律：$(A \cup B) \cup C = A \cup (B \cup C),\ (AB)C = A(BC)$。

3) 分配律：$(A \cup B)C = (AC) \cup (BC),\ (AB) \cup C = (A \cup C)(B \cup C)$。

4) De Morgan 定律：$(A \cup B)^c = A^c B^c,\ (AB)^c = A^c \cup B^c$.

多个、可列个事件的交、并运算

$$\bigcap_{i=1}^{n} A_i = A_1 A_2 \cdots A_n, \qquad \bigcap_{i=1}^{\infty} A_i = A_1 A_2 \cdots$$

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n, \qquad \bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$$

若 $A_1, A_2, \cdots$ 两两互斥，则记 $\sum_{i=1}^{n} A_i = \bigcup_{i=1}^{n} A_i,\ \sum_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} A_i$

分配律：$A(\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} (AA_i)$

De Morgan Th. $(\bigcap_{i=1}^{\infty} A_i)^c = \bigcup_{i=1}^{\infty} A_i^c, \qquad (\bigcup_{i=1}^{\infty} A_i)^c = \bigcap_{i=1}^{\infty} A_i^c$

极限

若 $A_1 \subset A_2 \subset \cdots$，则记 $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$；

若 $A_1 \supset A_2 \supset \cdots$，则记 $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$。

2. 概率的 Kolmogorov 公理化定义

对于一个随机试验，如何合理地定量描述各事件发生的可能性大小，这是概率论的基本任务。

定义 1.1 （σ 域、事件域）设 S 是一个样本空间，F 为由 S 的某些子集构成的一个非空集合类。若 F 满足下列三条性质：

a) $S \in \mathrm{F}$；

b) 若 $A \in \mathrm{F}$，则 $A^c \in \mathrm{F}$；

c) 若 $A_1, A_2, \cdots$ 是 F 中的可列个元素，则 $\bigcup_{i=1}^{\infty} A_i \in \mathrm{F}$。

则称 F 为 S 上的一个 σ 域（事件域）。

例 1.2

1) $S=\{s_1, s_2, \mathbf{L}, s_n\}$，F 是 S 的所有子集组成的集合类，则它是一个σ 域。

2) ¡ 上的 Borel 域 B_{i} ：包含了所有形如 $(-\infty, a]$ 的区间的最小σ 域，$a\in$ ¡；

¡^{k} 上的 Borel 域 $\mathrm{B}_{\mathrm{i}^k}$：包含了所有形如 $(-\infty, a_1]\times(-\infty, a_2]\times\mathbf{L}\times(-\infty, a_k]$ 的区域的最小σ 域，$(a_1, a_2, \mathbf{L}, a_k)\in$ ¡^{k}。

定义 1.2 （概率）对于一个给定的样本空间 S 及 S 上的一个事件域 F，定义在 F 上的实值函数 $P(\cdot)$ 若满足下列三条公理：

a) （非负性）$P(A)\geq 0,\ \forall A\in \mathrm{F}$；

b) （规范性）$P(S)=1$；

c) （可列可加性）若 $A_1, A_2, \mathbf{L}$ 是 F 中的可列个互不相容的事件，则

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

则称 $P(\cdot)$ 是 (S, F) 上的一个概率（函数），称 (S, F, P) 是一个概率空间。

注：

Ⅰ Kolmogorov 公理化定义只是明确了作为概率的数学模型应该满足的基本条件，并不涉及对概率的解释，也不涉及如何确定具体事件的概率值。

3. 概率的基本性质

(1) $P(\varnothing)=0$。

(2)（有限可加性）设 $A_i\in \mathrm{F}$，$i=1, \mathbf{L}, n$ 且两两互斥，则 $P(\sum_{i=1}^{n} A_i)=\sum_{i=1}^{n} P(A_i)$。

(3) $\forall A\in \mathrm{F}$， i) $P(A^c)=1-P(A)$； ii) $P(A)\leq 1$。

(4) 若 $A\subset B$，则 $P(A)\leq P(B)$。

(5) （减法公式）$P(B-A)=P(BA^c)=P(B)-P(BA)$。

(6) （加法公式）

$$P(A\cup B)=P(A)+P(B)-P(AB);$$
$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC);$$
$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(\bigcap_{i=1}^{n}A_i).$$

(7) （次可列可加性、Boole 不等式） $P(\bigcup_{i=1}^{\infty}A_i)\le\sum_{i=1}^{\infty}P(A_i)$。

（Bonferroni 不等式） $P(\bigcap_{i=1}^{n}A_i)\ge\sum_{i=1}^{n}P(A_i)-(n-1)$。

(8) 可列可加 $\Longleftrightarrow$ 有限可加 + $\varnothing$上连续 $\Longleftrightarrow$有限可加 + 下连续。

4. 确定概率的例子

(1) 有限样本空间。

设 $S=\{s_1,s_2,\cdots,s_n\}$，$\mathrm{F}=2^S$，设 $p_1,p_2,\cdots,p_n$ 是 n 个非负实数，且 $\sum_{i=1}^{n}p_i=1$，若对 $\forall A\in\mathrm{F}$，取 $P(A)=\sum_{s_i\in A}p_i$。则可以验证，这么定义的函数 $P(\cdot)$ 就是一个概率。$p_1,p_2,\cdots,p_n$取不同的值，则对应于 (S,F) 上不同的概率函数。

同理，也可确定无限可列样本空间上的概率。

(2) 古典概型 (Classical Scheme)。

是历史上研究得最早的概率模型。描述的是一类最简单的随机现象。概率的确定建立在经验事实的基础上（对基本事件作等概率的假定）。通过逻辑分析得到感兴趣事件的概率。

特点：

a) 样本空间有限，不妨记作 $S=\{s_1,s_2,\cdots,s_n\}$；

b) 假定每个基本事件 $\{s_i\}$，$i=1,\cdots,n$ 发生的概率相等，即

$$P(\{s_i\})=\frac{1}{n},\ i=1,\cdots,n。$$

c) 事件 A 的概率

$$P(A)=\frac{A\text{中包含的样本点数}}{S\text{中的样本点数}}。$$

容易验证，按上述方法得到的概率，符合 Kolmogorov 公理化定义。

例 1.3

1) 彩票中奖问题：

   买一注 35 选 7 的福利彩票，中各奖项的概率多大？

2) 生日问题：

   r (<365) 个人中至少有两人生日相同的概率多大？

3) 抽签问题：

   N 个签中有 A 个签有奖，各人依次抽签，第 t 人中奖的概率多大？

4) 抽样检验问题：

   N 个产品中有 M 个次品，其余合格。从中抽取 n 个，求恰好抽中 x 个次品的概率。分两种情况：有放回；无放回。

(3) 几何概型

a) 样本空间 S 是 $\mathbb{R}^k$ 中的一个区域，且有几何量度 $m(S)$；

b) “等可能性假定”：试验结果落在 S 中某一子区域 A 中的概率只与 A 的几何量度 $m(A)$ 有关，与 A 的形状、位置无关；

c) $P(A)=\dfrac{m(A)}{m(S)}$。

例 1.4

1) 会面问题

2) Buffon 投针问题

(4) 频率方法

§1.2 条件概率与事件的独立性

1. 条件概率

引进条件概率的作用：考察事件之间的关系；简化概率的计算。

定义 1.3 （条件概率）设 (S, F, P) 是一个概率空间，$B\in \mathrm{F}$，$P(B)>0$。对于 $\forall A\in \mathrm{F}$，称

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

为事件 B 发生条件下 A 发生的条件概率。

例 1.5 在一般人群中任取一人，有肺癌的概率极小。但若已知抽中的人有很长的吸烟史，则该人患肺癌的概率就大了。后者就是条件概率。例如，一个 1 万人的人群结构如下

	患肺癌 (A)	不患肺癌 (A^c)	合计
抽烟 (B)	80	1920	2000
不抽烟 (B^c)	8	7992	8000
合计	88	9912	10000

从中随机抽 1 人，此人有肺癌的概率为

$$P(A)=\frac{\binom{88}{1}}{\binom{10000}{1}}=0.0088,$$

若已知抽中的人抽烟，则其有肺癌的条件概率为

$$P(A\mid B)=\frac{\binom{80}{1}}{\binom{2000}{1}}=\frac{\binom{80}{1}\Big/\binom{10000}{1}}{\binom{2000}{1}\Big/\binom{10000}{1}}=0.04\text{。}$$

注：

- 条件概率可视为将样本空间 S 缩小到 $S_B=B$ 之后来研究各事件的概率。
- 对于给定的事件 B，$P(A\mid B),\ \forall A\in \mathrm{F}$ 形成了 (S, F) 上的一个新的概率函数，即满足 Kolmogorov 公理化定义的三条公理：
  a) （非负性）$P(A\mid B)\geq 0,\ \forall A\in \mathrm{F}$；
  b) （规范性）$P(S\mid B)=1$；

c) （可列可加性）若 $A_1, A_2, \mathbf{L}$ 是 F 中的可列个互不相容的事件，则

$$P(\sum_{i=1}^{\infty} A_i \mid B) = \sum_{i=1}^{\infty} P(A_i \mid B)。$$

而且，它具有概率的一切性质。

l 若 $B = S$，则 $\forall A \in \mathrm{F}$，$P(A \mid B) = P(A \mid S) = P(A)$，故无条件概率可视为特殊条件下的条件概率。

乘法公式：若 $A_1, A_2, \mathbf{L}, A_n$ 是 (S, F) 中的 n 个事件，且 $P(\prod_{i=1}^{n} A_i) > 0$，则

$$P(\prod_{i=1}^{n} A_i) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)\,\mathbf{L}\,P(A_n \mid A_1A_2\,\mathbf{L}\,A_{n-1})。$$

全概率公式：

若 $B_1, B_2, \mathbf{L}, B_n$ 互不相容，且 $P(B_i) > 0,\ i = 1, \mathbf{L}, n,\ A \subset \sum_{i=1}^{n} B_i$，则

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

例 1.6 （抽签问题）已知 100 个签中有 10 个有奖，每人依次抽一签。

(1) 求第三个人**首次**中奖的概率；

(2) 求第二个人中奖的概率。

Bayes 公式：

设 $B_1, B_2, \mathbf{L}, B_n$ 互不相容，且 $P(B_i) > 0,\ i = 1, \mathbf{L}, n$。$A \subset \sum_{i=1}^{n} B_i,\ P(A) > 0$。则

对 $\forall j \in \{1, 2, \mathbf{L}, n\}$，

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}。$$

例 1.7 （甲胎蛋白法检查肝癌）某地区肝癌发病率为 0.0004，当时用甲胎蛋白法检查肝癌，真阳性率为 99%，假阳性率为 0.1%。随机抽 1 人作检查，呈阳性。求此人患肝癌的概率。

解： 记 A 为此人患肝癌，B 为检查阳性。已知 $P(B|A)=0.99,\ P(B|A^c)=0.001$，

$P(A)=0.0004$，求 $P(A|B)$。由 Bayes 公式知

$$P(A|B)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$$
$$=\frac{0.99\times 0.0004}{0.99\times 0.0004+0.001\times 0.9996}=0.284.$$

2. 事件的独立性

独立性是概率论中一个重要概念，是用概率论的语言对经验直观的独立性概念给出的严格定义。

定义 1.4 （两事件的独立性）事件 A、B 若满足

$$P(AB)=P(A)P(B),$$

则称事件 A 与 B 相互独立。

注：

- 两事件独立意味着条件概率等于无条件概率，i.e. 若 A 与 B 相互独立，则

当 $P(B)>0$ 时，$P(A|B)=P(A)$；当 $P(A)>0$ 时，$P(B|A)=P(B)$。

- 概率为 0 或 1 的事件与任何事件独立；特别地，$S,\varnothing$ 与任何事件独立。

性质：

若 A 与 B 相互独立，则 A 与 B^c，A^c 与 B，A^c 与 B^c 也都相互独立。

定义 1.5 （多个事件的独立性）$A_1, A_2, \mathbf{L}, A_n$ 是 n 个事件，若对集合 $\{1,2,\mathbf{L},n\}$ 的任意一个至少含两个元素的子集 C 有

$$P(\mathop{\mathbf{I}}_{j\in C} A_j)=\prod_{j\in C}P(A_j),$$

则称事件 $A_1, A_2, \mathbf{L}, A_n$ 相互独立。

性质：

设事件 $A_1, A_2, \mathbf{L}, A_n$ 相互独立，则

a) 从中任取 m 个事件 $(2 \le m \le n)$ 也相互独立。

b) 从中任取 m 个事件 $(2 \le m \le n)$，对每个事件取余或不取余，得到的 m 个新事件也相互独立。

c) 将这 n 个事件任意分成 m 组，各组通过组内事件间的运算各自得到一个新事件，则这 m 个新事件相互独立。

d) 独立乘法公式：

$$P(\mathop{\mathbf{I}}_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)。$$

e) 独立加法公式：

$$P(\mathop{\mathbf{U}}_{i=1}^{n} A_i) = 1 - P(\mathop{\mathbf{I}}_{i=1}^{n} A_i^c) = 1 - \prod_{i=1}^{n} P(A_i^c)。$$

例 1.8（两两独立、但三事件不独立）设样本空间为

$$S = \begin{Bmatrix} (aaa), & (bbb), & (ccc), \\ (abc), & (bac), & (cba), \\ (acb), & (bca), & (cab) \end{Bmatrix}。$$

假定各基本事件等概率。记 A_i 为第 i 个位置上是字母 a，$i = 1,2,3$。则

$$P(A_i) = \frac{1}{3},\ i = 1,2,3; \qquad P(A_1A_2) = \frac{1}{9} = P(A_1)P(A_2),$$

$$P(A_1A_3) = \frac{1}{9} = P(A_1)P(A_3), \quad P(A_2A_3) = \frac{1}{9} = P(A_2)P(A_3).$$

说明 A_1, A_2, A_3 两两相互独立。但是 $P(A_1A_2A_3) = \frac{1}{9} \ne P(A_1)P(A_2)P(A_3)$，说明三事件不独立。

定义 1.6（试验的独立性）一个随机试验 E (an experiment) 由 n 个子试验 (subexperiment or trial) $E_1, E_2, \mathbf{L}, E_n$ 构成。设 E_i 的样本空间为 S_i，$i = 1, \mathbf{L}, n$，E 的样本空间为 $S = S_1 \times S_2 \times \mathbf{L} \times S_n$（笛卡尔乘积）。若对任意一组事件 $\{A^{(i)} \subset S_i : i = 1, \mathbf{L}, n\}$，其笛卡尔积 $A = A^{(1)} \times A^{(2)} \times \mathbf{L} \times A^{(n)}$ 为 E 的一个事件，都

有

$$P(A)=P\left(\bigcap_{i=1}^{n}A^{(i)}\right)=\prod_{i=1}^{n}P(A^{(i)})\text{，}$$

则称子试验 $E_1, E_2, \mathbf{L}, E_n$ 相互独立。

独立性概念的作用：

对于给定的概率模型，判断复杂事件间的独立性关系；

用于建模时以经验的独立性判断作为假定，简化概率模型。

第二章　随机变量

§2.1　随机变量及其分布

1. 随机变量 (random variable, r. v.)

在研究随机现象时，往往需要用数值记录观测结果，这就产生了随机变量的概念。

定义 2.1 （随机变量）定义在样本空间 S 上的实值函数 $X(s): S \to \mathbb{R}$，称为样本空间 S 上的随机变量。

注：

- 随机变量通常用大写英文字母 $X, Y, Z, \ldots$ 等表示，其取值常用小写字母 $x, y, z, \ldots$ 等表示。

- 有些样本空间本身就是某个实数的子集，观测结果本身就是一个随机变量。
  例如：
  - 掷一枚骰子出现的点数 X;
  - 观察某大卖场一天内来到的顾客数 X 及 营业额 Y;
  - 抽取某厂生产的某型号的电视机检测其寿命 T;

- 有些样本空间不是实数的子集，可根据研究需要定义适当的随机变量。
  例如：
  - 检验一个产品，看它是否合格。样本空间 S={合格，不合格}。可定义 r.v. X 为不合格品数。则

$X(s)$	s
0	合格
1	不合格

  - 观测投掷一枚硬币三次的结果。样本空间 S={HHH，HHT，HTH，THH，HTT，THT，TTH，TTT}。可定义 r.v. X 为正面数。则

$X(s)$	s
0	TTT
1	HTT，THT，TTH
2	HHT，HTH，THH
3	HHH

- 一个随机变量通常可视为考察随机现象的一个视窗。

导出概率分布

设 (S, F, P) 是一个概率空间，X 是定义在 S 上的一个随机变量。那么，由 X 可在实数空间 $(\mathbb{R}, \mathrm{B}_{\mathbb{R}})$ 上导出一个新的概率 $P_X(\cdot)$。$P_X(\cdot)$ 称为随机变量 X（导出）的概率分布，$\forall B \in \mathrm{B}_{\mathbb{R}}$，

$$P_X(B) = P(X \in B) = P(\{s : X(s) \in B\})。$$

注：

- 随机变量 X 的概率分布完全由 (S, F, P) 决定。
- 了解 X 的分布，至少可以了解 P 的某一个侧面的信息。
- 定义在不同概率空间上的随机变量可能有共同的分布，因此，抽象地研究随机变量的分布有助于找到共同的方法去研究同类随机现象。

2. 分布函数

定义 2.2 （分布函数）若 X 是一个随机变量，称函数

$$F(x) = P_X((-\infty, x]) = P(X \le x), \quad x \in \mathbb{R}$$

为 X 的（累积）分布函数 (d.f. / c.d.f.)。常记为 $X \sim F(\cdot)$。

注：

- 分布函数与概率分布一一对应。
- 引进分布函数，实质上是将 $(\mathbb{R}, \mathrm{B}_{\mathbb{R}}, P_X)$ 转化为一个实函数来研究。
- 通常用 F, G, H 等大写字母表示分布函数.

基本性质：

$F(\cdot)$ 是某个随机变量的分布函数，当且仅当 $F(\cdot)$ 同时满足下列三个条件：

a) （单调性）$\forall a < b,\ F(a) \le F(b)$；

b) （有界性）$\lim_{x \to -\infty} F(x) = 0,\ \lim_{x \to \infty} F(x) = 1$；

c) （右连续性）$\forall x_0 \in \mathbb{R},\ F(x_0) = \lim_{x \to x_0 +} F(x)$。

注：

- 分布函数的另一种定义为：$\tilde{F}(x) = P(X < x),\ x \in \mathbb{R}$。此时，基本性质 a) b) 不变，c) 变为“左连续”。

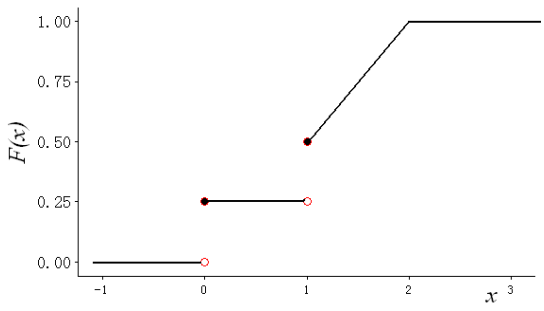
定义 2.3 （**同分布**）若随机变量 X 与 Y 的分布函数相同，则称 X 与 Y 同分布。

用分布函数可以计算哪些事件的概率？

- $P(X>a)=1-P(X\le a)=1-F(a)$
- $P(a<X\le b)=P(\{X\le b\}-\{X\le a\})=F(b)-F(a)$
- $P(X<a)=\lim\limits_{x\to a-}F(x)=F(a-0)$
- $P(X=a)=P(\{X\le a\}-\{X<a\})=F(a)-F(a-0)$
- $P(X\ge a)=1-P(X<a)=1-F(a-0)$
- $P(a\le X\le b)=F(b)-F(a-0)$
- $P(a\le X<b)=F(b-0)-F(a-0)$
- $P(a<X<b)=F(b-0)-F(a)$

例 2.1 设 r.v. X 具有如下分布函数

$$F(x)=\begin{cases}0, & x<0,\\ \frac{1}{4}, & 0\le x<1,\\ \frac{1}{2}x, & 1\le x<2,\\ 1, & 2\le x.\end{cases}$$



1) $P(X<1)=F(1-0)=\frac{1}{4}$。

2) $P(X=1)=F(1)-F(1-0)=\frac{1}{2}-\frac{1}{4}=\frac{1}{4}$。

3) $P(\frac{1}{2}\le X<\frac{3}{2})=F(\frac{3}{2}-0)-F(\frac{1}{2}-0)=\frac{1}{2}\frac{3}{2}-\frac{1}{4}=\frac{1}{2}$。

4) $P(X=\frac{3}{2})=F(\frac{3}{2})-F(\frac{3}{2}-0)=0$。

5) 求 c 使得 $P(X\le c)=\frac{2}{3}$。即求 c 使得 $F(c)=\frac{2}{3}$，解得 $c=\frac{4}{3}$。

6) 求 c 使得 $P(X\le c)=0.4$；求 c 使得 $P(X\le c)=0.25$。

定义 2.4 （分位点）设 r.v. X 的分布函数为 $F(x)$。对于给定的 $p\in(0,1)$，若 a_p 满足

$$F(a_p-0)\le p\le F(a_p),$$

则称 a_p 为 X 的 p 分位点 (quantile)。

一些常用的分位点及其特殊名称：

$a_{0.5}$ 中位数 (median)

$a_{0.25}$ 第一四分位数 (first quartile)， $a_{0.75}$ 第三四分位数 (third quartile)

$a_{0.1}, a_{0.2}, \mathbf{L}, a_{0.9}$ 第一～第九十分位数 (first decile ～ ninth decile)

3. 离散型随机变量与分布列

若一个随机变量最多只可能取可列个值，则称之为离散型随机变量。

定义 2.5 （分布列）若离散型随机变量 X 的可能取值为 $x_i, i=1,2,\mathbf{L}$，则称数列

$$p_i \triangleq P(X=x_i),\ i=1,2,\mathbf{L}$$

为 X 的概率分布列 (probability mass function/discrete probability density function)。

分布列与分布函数的关系：

- 离散型随机变量的分布函数必是阶梯函数，与分布列一一对应。
- 在定义 2.5 的条件下，有

$$p_i=F(x_i)-F(x_i-0),\ i=1,2,\mathbf{L}$$
$$F(a)=P(X\le a)=\sum_{x_i\le a}p_i,\ \forall a\in \text{¡}$$

- 根据分布列可计算离散型随机变量落在任意给定的实数区间内的概率。

分布列的基本性质：

一个数列 $\{p_i : i=1,2,\mathbf{L}\}$ 是某个离散型随机变量的分布列，当且仅当它同时满足下列两个条件：

a) （非负性） $p_i\ge 0,\ i=1,2,\mathbf{L}$；

b) （规范性） $\sum_{i=1}^{\infty}p_i=1$。

分布列的表示方法：

a) 通项公式； b) 表格形式； c) 条形图。

4. 连续型随机变量与密度函数

定义 2.6 （连续型随机变量）若存在非负函数 $p(x),\ x\in$ ¡ ，使得随机变量 X 的分布函数 $F(x)$ 可表示为

$$F(x)=\int_{-\infty}^{x}p(t)dt,\quad \forall x\in \text{¡}\ ,$$

则称 X 是连续型随机变量，称 $p(x)$ 为 X 的概率密度函数 (p.d.f.)。

密度函数的基本性质：

一个函数 $p(x)$ 是某个连续型随机变量的密度函数，当且仅当它同时满足下列两个条件：

a) （非负性） $p(x)\geq 0,\ \forall x\in$ ¡ ；

b) （规范性） $\int_{-\infty}^{\infty}p(x)dx=1$。

连续型随机变量的特点：

1)

$$\forall a<b,\ P(a<X\leq b)=F(b)-F(a)$$
$$=\int_{-\infty}^{b}p(t)dt-\int_{-\infty}^{a}p(t)dt=\int_{a}^{b}p(t)dt.$$

$$\forall\ A\in \mathrm{B}_{\text{¡}},\ P(X\in A)=\int_{A}p(t)dt\,.$$

2) $P(X=c)=0,\ \forall c\in$ ¡ 。

3) 分布函数处处连续。

$$P(a<X\leq b)=P(a<X<b)=P(a\leq X<b)=P(a\leq X\leq b)=\int_{a}^{b}p(x)dx\text{。}$$

4) 连续型随机变量的密度函数不唯一。改变密度函数在有限个点、可列个点（甚至更多个点）处的函数值，不改变分布函数。

5) 若 $p(x)$ 在 x 处连续，则有 $F'(x)=p(x)$。因此，若分布函数 $F(x)$ 处处可导，则可取 $F'(x)$ 作为密度函数。

密度函数的来源——直方图

§2.2 常用概率分布族

1. 离散型分布

(1) 二项分布 (Binomial Distribution)

独立地重复进行 n 次成功概率为 p (0<p<1) 的 Bernoulli 试验，记成功次数为 X。称 X ～ B(n, p)。

分布列为

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,\mathbf{L},n\ 。$$

分布特点：

- p<0.5 时分布右偏，p=0.5 时分布对称，p>0.5 时分布左偏。
- (对称性) 若 X～B(n, p)，则 Y=n－X～B(n, 1－p)。
- 若记 X_i 为第 i 次 Bernoulli 试验中的成功次数，则 $X=\sum_{i=1}^{n}X_i$ 。

(2) 超几何分布 (Hypergeometric Distribution)

来源于对有限总体的不放回抽样。设一袋子中共有 N 球，其中有 M 个白球。从中不放回地任取 n 个球，记抽中的白球数为 X。称 X ～ HG(n; N, M)。

分布列为

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=\max\{0,n-(N-M)\},\mathbf{L},\min\{M,n\}\ 。$$

当 $n=N$ 且 $n<M$ 时，不放回抽样与放回抽样非常接近，有

$$P(X=k)\approx\binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}\ 。$$

(3) Poisson 分布

分布列为

$$P(X=k)=\frac{l^k}{k!}e^{-l},\quad k=0,1,2,\mathbf{L}$$

其中 l 是大于 0 的参数。记作 $X\sim\mathrm{Poi}(l)$。

来源：二项分布概率的近似计算。当 n 较大，np 较小时，对于给定的较小的 k，

$$\binom{n}{k}p^k(1-p)^{n-k} \approx \frac{l^k}{k!}e^{-l}, \quad \text{其中 } l = np \text{。}$$

应用：

- Ø 用于对一定时间段内某类事件发生的次数建模。基本假定：记在充分小的时间段 h 内，这类事件发生的次数为 Y，则 1) $P(Y=1)$ 与 h 成比例；2) $P(Y \geq 2) = 0$；3) 在不交叉的时间段内该类事件发生的次数相互独立。
- Ø 用于对一定空间范围内某类事物出现的个数建模。
- Ø 例
  - ² 1875－1894 年间普鲁士军队中每年被马踢死的士兵数；
  - ² 给定时间段内一个放射源发出的 a 粒子数；
  - ² 二次大战中，在固定时间长度内，伦敦被炮击中的城区数；
  - ² 一分钟内到达某银行的人数；
  - ² 一定时间内通过某个十字路口的汽车数；
  - ² 单位布匹上的疵斑数 … …

(4) 几何分布 (Geometric Distribution)

独立地重复进行成功概率为 p 的 Bernoulli 试验，记首次成功时的试验次数为 X。称 $X \sim \text{Ge}(p)$。

分布列为

$$P(X=k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \mathbf{L}$$

特点：无记忆性。对于 k>0，s>=0,

$$P(X = k+s \mid X > s) = \frac{P(X=k+s)}{P(X>s)} = \frac{(1-p)^{k+s-1}p}{\sum_{l=s+1}^{\infty}(1-p)^{l-1}p}$$
$$= (1-p)^{k-1}p = P(X=k).$$

(5) 负二项分布 (Negative Binomial Distribution / Pascal Distribution)

独立地重复进行成功概率为 p 的 Bernoulli 试验，记第 r 次成功时的试验次数为 X。称 $X \sim \text{NB}(r, p)$。

分布列为

$$P(X=k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \quad k = r, r+1, \mathbf{L}$$

若记第 i 1 次成功之后至第 i 次成功时的试验次数为 X_i，$i = 1, 2, \mathbf{L}, r$，则 $X_1, \mathbf{L}, X_r$ i.i.d. $\text{Ge}(p)$，且 $X = \sum_{i=1}^{r} X_i$。

2. 常用连续型分布

1) 均匀分布 (Uniform Distribution)

若 r.v. X 具有 p.d.f.

$$p(x)=\frac{1}{b-a}I_{[a,b]}(x)=\begin{cases}\dfrac{1}{b-a}, & \text{if } x\in[a,b],\\ 0, & \text{else.}\end{cases}$$

则称 X ～ U[a,b].

d.f. 为

$$F(x)=\int_{-\infty}^{x}p(t)dt=\begin{cases}0, & x\le a,\\ \dfrac{x-a}{b-a}, & a<x\le b,\\ 1, & x>b.\end{cases}$$

用于描述一类几何概型。

例如，Bertrand 悖论。在单位圆内任取一条弦，求弦长大于内接等边三角形的边长的概率。按照不同方法取弦，则会得到不同的概率值。

a) 假定在一条直径上按均匀分布随机取一点，过该点作垂直于该直径的弦。则所求概率为 1/2。

b) 在圆上任取一点，按 U[0, π] 任取一个角度作为弦切角大小作弦，则所求概率为 1/3。

c) 在圆内任取一点作为弦的中点。所求概率等于？

2) Gamma 分布

常用作对“寿命”建模。

若 r.v. X 具有 p.d.f.

$$p(x;a,b)=\frac{1}{\Gamma(a)b^{a}}x^{a-1}e^{-x/b}I_{(0,\infty)}(x),$$

其中 $a>0$ 称为形状参数 (shape parameter)，$b>0$ 称为尺度参数 (scale parameter)。

则称 $X\sim\text{Gamma}(a,b)$.

特例：

- 卡方分布 (Chi square distribution)。当 $b=2$ 时，令 $a=\frac{r}{2}$，则 gamma 分布的 p.d.f. 变为

$$p(x;r)=\frac{1}{\Gamma(\frac{r}{2})2^{r/2}}x^{r/2-1}e^{-x/2}I_{(0,\infty)}(x),$$

称为自由度为 r 的卡方分布，记为 $X\sim c^2(r)$。

- 指数分布 (Exponential distribution)。当 $a=1$ 时，gamma 分布的 p.d.f. 变为

$$p(x;b)=\frac{1}{b}e^{-x/b}I_{(0,\infty)}(x),$$

称为 $X\sim \mathrm{Exp}(b)$。

指数分布具有无记忆性，即对于 $\forall s>0,\ \forall t\geq 0$，有

$$P(X>s+t\mid X>t)=P(X>s).$$

Gamma 分布与 Poisson 分布的关系

若 $X\sim \mathrm{Gamma}(a,b)$，$a$ 是一个正整数，那么，对于 $\forall x>0$ 有

$$P(X\leq x)=P(Y\geq a),$$

其中 $Y\sim \mathrm{Poi}(x/b)$.

3) 正态分布 (Normal Distribution / Gauss Distribution)

p.d.f.

$$p(x;m,s^2)=\frac{1}{\sqrt{2p}\,s}\exp\{-\frac{1}{2s^2}(x-m)^2\},\quad x\in \mathbb{R}.$$

其中 $m\in\mathbb{R}$，$s>0$ 为参数。记作 $X\sim N(m,s^2)$。

特点：

- p.d.f. 关于 $x=m$ 对称；函数值在 $x=m$ 处最大，有界；$x\to\pm\infty$ 时，$p(x)\to 0$。
- $x=m\pm s$ 是 p.d.f. 的两个拐点。
- 固定 s，改变 m 的值，p.d.f. 曲线形状不变，位置随 m 平移；固定 m 改变 s，则 p.d.f. 中心位置不变，形状随 s 变瘦高或矮胖。

标准正态分布：

- $m=0,\ s^2=1$ 的正态分布为标准正态分布，i.e. N(0,1)。
- 记 p.d.f. 为 $f(x)$，d.f. 为 $\Phi(x)$。

² 与 $N\left(m, s^2\right)$ 的分布函数 $F(x)$、密度函数 $f(x)$之间的关系：

$$F(x)=\Phi(\frac{x-m}{s}),\qquad f(x)=\frac{1}{s}f(\frac{x-m}{s}).$$

4) Beta 分布

p.d.f. 为

$$p(x;a,b)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}I_{(0,1)}(x),$$

其中 $a>0$， $b>0$ 为参数。

与二项分布的关系：

设 $X\sim \text{Beta}(a,b)$，其中 a,b 都是正整数且 $b=n-a+1$，则对 $\forall\, x\in(0,1)$ 有

$$F(x)=P(Y\geq a),$$

其中 $F(x)$ 为 X 的 d.f.，$Y\sim B(n,x)$.

5) Cauchy 分布

p.d.f. 为

$$p(x;q)=\frac{1}{p[1+(x-q)^2]},\ x\in(-\infty,+\infty),$$

其中 $q\in(-\infty,\infty)$ 为参数。

第三章　随机向量

§3.1　随机向量及其联合分布

1. 随机向量

定义 3.1（**随机向量**）定义在样本空间 S 上的 k 个随机变量组成的 k 维向量 $(X_1(s), \mathbf{L}, X_k(s))$ 称为 k 维随机向量 (random vector, r. vec.)。

定义 3.2（**联合分布函数**）设 $\underset{\sim}{X} = (X_1, \mathbf{L}\, X_k)$ 是 k 维 r. vec.，则称 k 元实值函数

$$F(x_1, \mathbf{L}, x_k) = P(X_1 \le x_1, \mathbf{L}, X_k \le x_k), \quad (x_1, \mathbf{L}, x_k) \in \mathbf{R}^k$$

为 r. vec. $\underset{\sim}{X}$ 的联合分布函数 (j.d.f.)。

注：

- 定义在样本空间 S 上的 k 维随机向量 $\underset{\sim}{X}$，就是一个 $S \to \mathbf{R}^k$ 的映射。通过 $\underset{\sim}{X}$ 可由概率空间 $(S, \mathbf{F}, P)$ 在 $(\mathbf{R}^k, \mathbf{B}_{\mathbf{R}^k})$ 上导出一个新的概率 P_X，称为 $\underset{\sim}{X}$（导出）的概率分布。$\forall B \in \mathbf{B}_{\mathbf{R}^k}$，

$$P_X(B) = P(\underset{\sim}{X} \in B) = P(\{s : \underset{\sim}{X}(s) \in B\}),$$

其中 $\mathbf{B}_{\mathbf{R}^k}$ 是 k 维 Borel 域，它是包含所有 k 维“长方体”的最小 sigma 域。

- $\underset{\sim}{X}$ 的概率分布与其联合分布函数一一对应。

- $\{X_1 \le x_1, \mathbf{L}, X_k \le x_k\} = \prod_{i=1}^{k} \{s : X_i(s) \le x_i\}$

- j.d.f. 是理论研究的重要工具，但一般难以用于计算概率。

联合分布函数的基本性质：

若 $F(\cdot)$ 是某个 k 维随机向量的联合分布函数，则 $F(\cdot)$ 必满足：

a) (单调性)

$\forall x_{j1} < x_{j2}$，$F(x_1, \mathbf{L}, x_{j-1}, x_{j1}, x_{j+1}, \mathbf{L}, x_k) \le F(x_1, \mathbf{L}, x_{j-1}, x_{j2}, x_{j+1}, \mathbf{L}, x_k)$;

b) (有界性)

$$F(x_1,\mathbf{L},-\infty,\mathbf{L},x_k)=\lim_{x_j\to-\infty}F(x_1,\mathbf{L},x_{j-1},x_j,x_{j+1},\mathbf{L},x_k)=0,\ j=1,\mathbf{L},k,$$

$$F(+\infty,\mathbf{L},+\infty)=1.$$

c) (右连续性)

$\forall \underset{\sim}{x}\in \mathbf{R}^k$, $F(x_1,\mathbf{L},x_j+0,\mathbf{L},x_k)=F(x_1,\mathbf{L},x_j,\mathbf{L},x_k),\ j=1,\mathbf{L},k$。

这三条是 j.d.f. 的必要条件，但不充分。对于 $k=2$ 的情形，除 a)~c) 外二元函数 F(.) 还需满足

d) $\forall a_1<b_1, a_2<b_2$， $F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2)\ge 0$

才能成为二元 j.d.f.。

例 3.1 二元函数

$$G(x,y)=\begin{cases}0, & x+y<0,\\ 1, & x+y\ge 0.\end{cases}$$

满足二维 j.d.f. 的性质 a)~c)，但不满足 d)，故不是一个 j.d.f.。

2. 离散型随机向量

定义 3.3 全部由离散型随机变量组成的随机向量称为离散型随机向量。

定义 3.4 设 k 维离散型随机向量 $\underset{\sim}{X}$ 的可能取值范围为 D（必是一个可列集），称

$$\{p_{x_1,\mathbf{L},x_k}\triangleq P(X_1=x_1,\mathbf{L},X_k=x_k):\ (x_1,\mathbf{L},x_k)\in D\}$$

为 $\underset{\sim}{X}$ 的联合分布列。

注：

- 离散型随机向量的性质完全由其联合分布列决定，分布列与 j.d.f.一一对应。
- 联合分布列的基本性质：

  a) 非负性：$p_{x_1,\mathbf{L},x_k}\ge 0,\quad \forall(x_1,\mathbf{L}\,x_k)\in D$；

  b) 规范性：$\sum_{(x_1,\mathbf{L},x_k)\in D}p_{x_1,\mathbf{L},x_k}=1$。

- $\forall A\in \mathbf{B}^k$， $P(\underset{\sim}{X}\in A)=\sum_{(x_1,\mathbf{L},x_k)\in A\mathbf{I}D}p_{x_1,\mathbf{L},x_k}$。
- 二维联合分布列也常用表格形式表示。

例 3.2 从 1、2、3、4 中随机取一个数记为 X，再从 1~X 中随机取一个数记为 Y。求：(1) (X,Y) 的联合分布列；(2) P(X=Y)。

解：(X, Y) 的可能取值范围为 $D=\{(i,j):\ j=1,\mathbf{L},i,\ i=1,\mathbf{L},4\}$。$\forall(i,j)\in D$，

$$P(X=i,Y=j)=P(Y=j\mid X=i)P(X=i)=\frac{1}{i}\cdot\frac{1}{4}。$$

联合分布列可用表格形式表示为：

X \ Y	1	2	3	4
1	1/4	0	0	0
2	1/8	1/8	0	0
3	1/12	1/12	1/12	0
4	1/16	1/16	1/16	1/16

$$P(X=Y)=P\left(\sum_{i=1}^{4}\{X=i,Y=i\}\right)=\frac{1}{4}+\frac{1}{8}+\frac{1}{12}+\frac{1}{16}=\frac{25}{48}=0.5208。$$

(X,Y) 的联合分布函数为

$$F(x,y)=\begin{cases}0, & y<1 \text{ or } x<1,\\ 1/4, & 1\le y,\ 1\le x<2,\\ 1/4+1/8, & 1\le y<2,\ 2\le x<3,\\ 1/4+2/8, & 2\le y,\ 2\le x<3,\\ 1/4+1/8+1/12, & 1\le y<2,\ 3\le x<4,\\ 1/4+2/8+2/12, & 2\le y<3,\ 3\le x<4,\\ 1/4+2/8+3/12, & 3\le y,\ 3\le x<4,\\ 1/4+1/8+1/12+1/16, & 1\le y<2,\ 4\le x,\\ 1/4+2/8+2/12+2/16, & 2\le y<3,\ 4\le x,\\ 1/4+2/8+3/12+3/16, & 3\le y<4,\ 4\le x,\\ 1, & 4\le y,\ 4\le x,\end{cases}$$

例 3.3 多项分布(Multinomial Distribution) 一次试验的结果可分为 $A_1,\mathbf{L},A_r$ r 个互斥的类；一次试验的结果落入 A_j 类的概率为 $P(A_j)=p_j>0,\ j=1,\mathbf{L},r$，其中 $p_1+p_2+\mathbf{L}+p_r=1$。独立地重复 n 次试验，记 A_j 类结果出现的次数为 $X_j,\ j=1,\mathbf{L},r$。则称随机向量 $(X_1,X_2,\mathbf{L},X_{r-1})\sim MN(n;p_1,p_2,\mathbf{L},p_r)$。

联合分布列为

$$\begin{aligned}P(X_1=x_1,\mathbf{L},X_{r-1}=x_{r-1})&=\binom{n}{x_1}\binom{n-x_1}{x_2}\mathbf{L}\binom{x_r}{x_r}p_1^{x_1}p_2^{x_2}\ \mathbf{L}\ p_r^{x_r}\\&=\frac{n!}{x_1!x_2!\mathbf{L}\,x_r!}p_1^{x_1}p_2^{x_2}\ \mathbf{L}\ p_r^{x_r},\end{aligned}$$

其中 $x_j\in\{0,1,\mathbf{L},n\},\ j=1,\mathbf{L},r$，且 $\sum_{j=1}^{r}x_j=n$。

3. 连续型随机向量

定义 3.5 设随机向量 $\underset{\sim}{X}=(X_1,\mathbf{L},X_k)$ 的 j.d.f. 为 $F(x_1,\mathbf{L},x_k)$。若存在非负实函数 $p(x_1,\mathbf{L},x_k)$ 使得

$$F(x_1,\mathbf{L},x_k)=\int_{-\infty}^{x_1}\mathbf{L}\int_{-\infty}^{x_k}p(t_1,\mathbf{L},t_k)dt_1\,\mathbf{L}\,dt_k,\quad \forall(x_1,\mathbf{L},x_k)\in \mathbf{¡}^k,$$

则称 $\underset{\sim}{X}$ 为 k 维连续性随机向量，称 $p(x_1,\mathbf{L},x_k)$ 为 $\underset{\sim}{X}$ 的联合密度函数 (j.p.d.f.)。

联合密度函数的基本性质：

(1) 非负性：$p(x_1,\mathbf{L},x_k)\ge 0,\quad \forall(x_1,\mathbf{L}\,x_k)\in \mathbf{¡}^k$；

(2) 规范性：$\int_{-\infty}^{\infty}\mathbf{L}\int_{-\infty}^{\infty}p(x_1,\mathbf{L},x_k)dx_1\,\mathbf{L}\,dx_k=1$。

其他性质：

l $\forall A\in \mathrm{B}_{\mathbf{¡}^k}$，$P(\underset{\sim}{X}\in A)=\int_A\mathbf{L}\int p(t_1,\mathbf{L},t_k)dt_1\,\mathbf{L}\,dt_k$。有关连续型随机变量的事件概率可通过联合密度函数的重积分来计算。

l 若 $p(x_1,\mathbf{L},x_k)$ 在 $(x_1,\mathbf{L}\,x_k)$ 处连续，则

$$\frac{\partial^k F(x_1,\mathbf{L},x_k)}{\partial x_1\,\mathbf{L}\,\partial x_k}=p(x_1,\mathbf{L},x_k)。$$

l 连续型随机向量的性质完全由其 jpdf 决定，jpdf 与 j.d.f.一一对应。

例 3.4 设 (*X*,*Y*) 的 j.p.d.f. 为

$$p(x,y)=\begin{cases}Ae^{-2x-y}, & x>0,y>0,\\ 0, & \text{otherwise.}\end{cases}$$

求：(1) 参数 A；(2) (*X*,*Y*)的 j.d.f.；(3) P(*X*<1, *Y*>1)；(4) P(*X*>*Y*)。

解：(1) A 应使联合密度函数满足规范性，即

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)dxdy=\int_0^{\infty}\int_0^{\infty}Ae^{-2x-y}dxdy$$
$$=A\cdot\left[-\frac{1}{2}e^{-2x}\Big|_0^{\infty}\right]\left[-e^{-y}\Big|_0^{\infty}\right]=\frac{1}{2}A,$$

因此，$A=2$。

(2) $\forall (x,y)\in \mathbb{R}^2$,

$$\begin{aligned} F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)dvdu \\ &= \begin{cases} 0, & x\le 0 \text{ or } y\le 0, \\ 2\int_0^x e^{-2u}du\cdot\int_0^y e^{-v}dv, & x>0 \text{ and } y>0. \end{cases} \\ &= \begin{cases} 0, & x\le 0 \text{ or } y\le 0, \\ \left(1-e^{-2x}\right)\left(1-e^{-y}\right), & x>0 \text{ and } y>0. \end{cases} \end{aligned}$$

(3)

$$\begin{aligned} P(X<1, Y>1) &= \iint\limits_{\{(x,y):x<1,y>1\}} p(x,y)dxdy \\ &= \int_0^1\int_1^\infty 2e^{-2x-y}dxdy = (1-e^{-2})e^{-1}. \end{aligned}$$

(4)

$$\begin{aligned} P(X>Y) &= \iint\limits_{\{(x,y):x>y\}} p(x,y)dxdy \\ &= \int_0^\infty\left(\int_0^x 2e^{-2x-y}dy\right)dx \\ &= \int_0^\infty 2e^{-2x}(1-e^{-x})dx \\ &= 1-\frac{2}{3}=\frac{1}{3}. \end{aligned}$$

常用连续型多元分布

(1) **均匀分布**。设 D 是 $\mathbb{R}^k$ 中的子区域，“体积”为 V>0。若 r. vec. $\underset{\sim}{X}=(X_1,\mathbf{L},X_k)$ 具有 j.p.d.f.

$$p(x_1,\mathbf{L},x_k)=\frac{1}{V}I_D(x_1,\mathbf{L},x_k),$$

则称 $\underset{\sim}{X}\sim \mathrm{U}(D)$。

(2) **多元正态分布**。设 $\underset{\sim}{m}=(m_1,\mathbf{L},m_k)'$ 是 k 维实向量，Σ 是 k 阶正定矩阵。称具有 j.p.d.f.

$$p(\underset{\sim}{x})=(2p)^{-k/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{m})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{m})\right],\quad \underset{\sim}{x}\in\mathbb{R}^k$$

的 r. vec. $\underset{\sim}{X}\sim N_k(\underset{\sim}{m},\Sigma)$。

§3.2 边际分布

设 $\underset{\sim}{X}=(X_1,\mathbf{L},X_k)$ 是 k 维 r. vec.， $\underset{\sim}{Y}$ 是 $\underset{\sim}{X}$ 的一个 m 维子向量， $m\in\{1,2,\mathbf{L},k-1\}$，

则称 $\underset{\sim}{Y}$ 的分布是 $\underset{\sim}{X}$ 的边际分布 (marginal distribution)。

描述边际分布的工具：边际分布函数；边际分布列；边际密度函数。

1. 边际分布函数 (m.d.f.)

l k=2。设 (X_1,X_2) 的 j.d.f. 为 $F(x_1,x_2)$，则

X_1 的 m.d.f 为

$$\begin{aligned}F_1(x_1)&=P(X_1\le x_1)\\&=P(X_1\le x_1,X_2<\infty)\\&=\lim_{x_2\to\infty}F(x_1,x_2)\triangleq F(x_1,\infty),\ x_1\in\ ¡.\end{aligned}$$

同理， X_2 的 m.d.f 为

$$F_2(x_2)=F(\infty,x_2),\ x_2\in\ ¡.$$

l k=3。设 (X_1,X_2,X_3) 的 j.d.f. 为 $F(x_1,x_2,x_3)$，则

X_1 的 m.d.f： $F_1(x)=F(x,\infty,\infty),\ x\in\ ¡.$

X_2 的 m.d.f： $F_2(x)=F(\infty,x,\infty),\ x\in\ ¡.$

X_3 的 m.d.f： $F_3(x)=F(\infty,\infty,x),\ x\in\ ¡.$

(X_1,X_2) 的 m.d.f： $F_{12}(x_1,x_2)=F(x_1,x_2,\infty),\ (x_1,x_2)\in\ ¡^2.$

2. 边际分布列

设离散型随机向量 $\underset{\sim}{X}=(X_1,\mathbf{L},X_k)$ 的联合分布列为

$$\{p_{x_1,\mathbf{L},x_k}\triangleq P(X_1=x_1,\mathbf{L},X_k=x_k):\ (x_1,\mathbf{L},x_k)\in D\},$$

则 X_j 的边际分布列为

$$p_j(x_j)=P(X_j=x_j)=\sum_{x_1}\mathbf{L}\sum_{x_{j-1}}\sum_{x_{j+1}}\mathbf{L}\sum_{x_k}P(X_1=x_1,\mathbf{L},X_k=x_k)。$$

例 3.5 从一个装有 2 个白球、3 个黑球的袋中随机地取球两次，每次一个。记第 *i* 次取得的白球数为 X_i，$i=1,2$。求 (X_1, X_2) 的联合分布列及边际分布列。

(1) 有放回地取。

X_1 \ X_2	0	1	X_1 的边际分布列
0	$(3/5)^2$	(3/5)(2/5)	3/5
1	(3/5)(2/5)	$(2/5)^2$	2/5
X_2 的边际分布列	3/5	2/5	

(2) 无放回地取。

X_1 \ X_2	0	1	X_1 的边际分布列
0	(3/5)(2/4)	(3/5)(2/4)	3/5
1	(3/4)(2/5)	(1/4)(2/5)	2/5
X_2 的边际分布列	3/5	2/5	

注：这两种情况下，两个分量的边际分布相同，但联合分布不同！所以，**边际分布不能确定联合分布。**

例 3.6 $(X_1, X_2, \mathbf{L}, X_{r-1}) \sim MN(n; p_1, p_2, \mathbf{L}, p_r)$。求其边际分布列。

3. 边际密度函数

k=2。设 (X_1, X_2) 的 j.d.f. 为 $F(x_1, x_2)$，jpdf 为 $p(x_1, x_2)$。则

X_1 的 mdf 为

$$\begin{aligned} F_1(x) &= P(X_1 \le x, X_2 < \infty) \\ &= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} p(t_1, t_2) dt_2 \right) dt_1, \ x \in \text{¡}. \end{aligned}$$

因此，

$$f_1(x) = \int_{-\infty}^{\infty} p(x, v) dv, \ x \in \text{¡}$$

是 X_1 的 mpdf。

同理，

$$f_2(x) = \int_{-\infty}^{\infty} p(u, x) du, \ x \in \text{¡}$$

是 X_2 的 mpdf。

例 3.7 设 (*X*,*Y*) 具有 j.p.d.f.

$$p(x,y)=\begin{cases}1, & 0<x<1,\ |y|<x,\\ 0, & \text{otherwise.}\end{cases}$$

求：(1) 边际密度函数; (2) P(*Y*>1/2)。

例 3.8 设 $\underset{\sim}{X}\sim N_k(\underset{\sim}{m},\Sigma)$，求 $\underset{\sim}{X}^{(1)}=(X_1,\mathbf{L}\ X_r),\ r<k$ 的边际密度。

§3.3 条件分布

条件分布是条件概率概念的推广，用于描述随机变量之间的关系。设 $(\underset{\sim}{X},\underset{\sim}{Y})$ 是 k 维随机向量，在给定 $\underset{\sim}{Y}\in B$ 的条件下，m 维子向量 $\underset{\sim}{X}$ 的条件概率分布为

$$P(\underset{\sim}{X}\in A\,|\,\underset{\sim}{Y}\in B)=\frac{P(\underset{\sim}{X}\in A,\underset{\sim}{Y}\in B)}{P(\underset{\sim}{Y}\in B)},\ A\in B_{\mathrm{i}^m}。$$

1. 离散型随机向量

k=2 的情形。设 (X_1,X_2) 的联合分布列为

$$\{p_{ij}\triangleq P(X_1=x_{1i},X_2=x_{2j}):\ i=1,\mathbf{L},n_1,\ j=1,\mathbf{L},n_2\}。$$

若 $p_{i\cdot}=P(X_1=x_{1i})>0$，则在给定 $X_1=x_{1i}$ 的条件下，事件 $\{X_2=x_{2j}\}$ 的条件概率为

$$P(X_2=x_{2j}\,|\,X_1=x_{1i})=\frac{p_{ij}}{p_{i\cdot}}\triangleq p_{j|i}。$$

$\{p_{j|i},\ j=1,\mathbf{L},n_2\}$ 为给定 $X_1=x_{1i}$ 条件下 X_2 的条件分布列。

同理，$\{p_{i|j}\triangleq\frac{p_{ij}}{p_{\cdot j}},\ i=1,\mathbf{L},n_1\}$ 为给定 $X_2=x_{2j}$ 条件下 X_1 的条件分布列。

续例 3.5 (1) 有放回情形。

$$P(X_1=i\,|\,X_2=0)=\frac{P(X_1=i,\,X_2=0)}{P(X_2=0)}=\begin{cases}3/5, & i=0,\\ 2/5, & i=1.\end{cases}$$

$$P(X_1=i\,|\,X_2=1)=\frac{P(X_1=i,\,X_2=1)}{P(X_2=1)}=\begin{cases}3/5, & i=0,\\ 2/5, & i=1.\end{cases}$$

可见，已知 X_2 取值的情况下， X_1 的条件分布都相同，且都与其边际分布相同。这说明 X_1 与 X_2 之间没有关系。

(2) 无放回情形。

X_1 \ X_2	0	1	X_1 的边际分布列	X_2=0 时 X_1 的条件分布	X_2=1 时 X_1 的条件分布
0	(3/5)(2/4)	(3/5)(2/4)	3/5	1/2	3/4
1	(3/4)(2/5)	(1/4)(2/5)	2/5	1/2	1/4
X_2 的边际分布列	3/5	2/5			

可见，已知 X_1 的条件分布随 X_2 取值的不同而不同。这说明两者之间有关系。

2. 连续型随机向量

对于连续型分布的 r. vec. $(\underset{\sim}{X},\underset{\sim}{Y})$，若 $P(\underset{\sim}{Y}\in B)=0$，则给定 $\underset{\sim}{Y}\in B$ 条件下 $\underset{\sim}{X}$ 条件分布无法直接由条件概率来定义。但可以借助于极限的概念来处理。

以二维连续型随机向量 (X,Y) 的情况为例。若给定 y，对于 $\forall e>0$，$P(y-e<Y\le y+e)>0$，且若 $\forall x\in \mathbb{R}$，下列极限存在

$$\lim_{e\to 0+} P(X\le x\,|\,y-e<Y\le y+e),$$

则称此极限形成的函数为**给定 $Y=y$ 条件下 X 的条件分布函数**，记为 $F_{X|Y}(x\,|\,y)$。若存在非负 $f_{X|Y}(x\,|\,y)$，使得

$$F_{X|Y}(x\,|\,y)=\int_{-\infty}^{x} f_{X|Y}(u\,|\,y)du,\ \forall x\in \mathbb{R},$$

则称 $f_{X|Y}(x\,|\,y)$ 为**给定 $Y=y$ 条件下 X 的条件密度函数**。

定理 3.1 设 $(\underset{\sim}{X},\underset{\sim}{Y})$ 是 k 维连续型随机向量，具有 jpdf $f(\underset{\sim}{x},\underset{\sim}{y})$。$\underset{\sim}{Y}$ 是 $k-m$ 维子向量，$f_Y(\underset{\sim}{y})$ 是其边际密度。若 $f_Y(\underset{\sim}{y})>0$，则给定 $\underset{\sim}{Y}=\underset{\sim}{y}$ 条件下 $\underset{\sim}{X}$ 的条件密度函数为

$$f_{\underset{\sim}{X}|\underset{\sim}{Y}}(\underset{\sim}{x}\,|\,\underset{\sim}{y})=\frac{f(\underset{\sim}{x},\underset{\sim}{y})}{f_{\underset{\sim}{Y}}(\underset{\sim}{y})},\quad \underset{\sim}{x}\in \mathbb{R}^m。$$

注：由**定理 3.1** 知，jpdf $f(\underset{\sim}{x},\underset{\sim}{y})$ 与 条件 pdf 有如下关系

$$f(x,y)=\begin{cases} f_X(x)f_{Y|X}(y\mid x), & \text{if } f_X(x)>0,\\ 0, & \text{else.}\end{cases}=\begin{cases} f_Y(y)f_{X|Y}(x\mid y), & \text{if } f_Y(y)>0,\\ 0, & \text{else.}\end{cases}$$

例 3.9 设 (X,Y) 服从圆 $\{(x,y):\ x^2+y^2\le 1\}$ 内的均匀分布，求条件密度。

解：(X,Y) 的 jpdf 为 $f(x,y)=\dfrac{1}{p}I(x^2+y^2\le 1)$。$X$ 的 mpdf 为

$$\begin{aligned} f_X(x)&=\int_{-\infty}^{\infty}f(x,y)dy\\ &=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{p}dy\cdot I_{(-1,1)}(x)=\frac{2}{p}\sqrt{1-x^2}\,I_{(-1,1)}(x).\end{aligned}$$

故 $x\in(-1,1)$ 时 $f_X(x)>0$。给定 $X=x$ $(x\in(-1,1))$ 条件下 Y 的条件密度函数为

$$f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{2\sqrt{1-x^2}}I_{(-\sqrt{1-x^2},\sqrt{1-x^2})}(y).$$

同理可得，给定 $Y=y$ $(y\in(-1,1))$ 条件下 X 的条件密度函数为

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}=\frac{1}{2\sqrt{1-y^2}}I_{(-\sqrt{1-y^2},\sqrt{1-y^2})}(x).$$

可见，mpdf $f_X(x)$ 与条件 pdf $f_{X|Y}(x\mid y)$ 是不同的。这说明 X 与 Y 是有关系的。

例 3.10 设 $(\underset{\sim}{X},\underset{\sim}{Y})\sim N_k(\underset{\sim}{m},\Sigma)$，其中 $\underset{\sim}{X}$ 是 m 维子向量。求给定 $\underset{\sim}{Y}=\underset{\sim}{y}$ 条件下 $\underset{\sim}{X}$ 的条件密度。

§3.4 随机变量的独立性

随机变量之间的独立性，也是随机事件独立性概念的推广。

定义 3.6 设 $(\underset{\sim}{X}_1,\mathbf{L},\underset{\sim}{X}_r)$ 是 k 维 r. vec.，其中 $\underset{\sim}{X}_j$ 是 k_j 维子向量，$\sum_{j=1}^{r}k_j=k$。若有

$$P\left(\underset{\sim}{X}_j\le\underset{\sim}{x}_j,\ j=1,\mathbf{L},r\right)=\prod_{j=1}^{r}P(\underset{\sim}{X}_j\le\underset{\sim}{x}_j),\quad \forall(\underset{\sim}{x}_1,\mathbf{L},\underset{\sim}{x}_r)\in \mathbb{R}^k,$$

则称子向量 $\underset{\sim}{X}_1,\mathbf{L},\underset{\sim}{X}_r$ 之间相互独立。

注：若$(X_1,\mathbf{L},X_r)$的 jdf 为 $F(x_1,\mathbf{L},x_r)$，则 $X_1,\mathbf{L},X_r$ 之间相互独立等价于

$$F(x_1,\mathbf{L},x_r)=\prod_{j=1}^{r}F_j(x_j),\quad \forall(x_1,\mathbf{L},x_r)\in \mathbf{R}^k,$$

其中 $F_j(\cdot)$ 是 X_j 的 mdf。

定理 3.2 随机向量 $X_1,\mathbf{L},X_r$ 之间相互独立等价于

$$P\left(\bigcap_{j=1}^{r}\{X_j\in A_j\}\right)=\prod_{j=1}^{r}P(X_j\in A_j),\quad \forall\, A_1\in\mathcal{B}_{\mathbf{R}^{k_1}},\mathbf{L},A_r\in\mathcal{B}_{\mathbf{R}^{k_r}}.$$

注：

- 对于离散型随机向量，$X_1,\mathbf{L},X_r$ 之间相互独立等价于

$$P\left(X_j=x_j,\ j=1,\mathbf{L},r\right)=\prod_{j=1}^{r}P(X_j=x_j),\quad \forall(x_1,\mathbf{L},x_r)\in D\text{。}$$

- 对于连续型随机向量，若 $p_j(\cdot)$ 是 X_j 的 mpdf，则 $X_1,\mathbf{L},X_r$ 之间相互独立等价于

$$p(x_1,\mathbf{L},x_r)=\prod_{j=1}^{r}p_j(x_j),\quad \forall(x_1,\mathbf{L},x_r)\in \mathbf{R}^k$$

是$(X_1,\mathbf{L},X_r)$的 jpdf。

- 对于随机向量(X,Y)，若对于任意使 $p_Y(y)>0$ 的 y，在给定 $Y=y$ 条件下，X 的条件分布都与 y 无关，则 X, Y 相互独立。

性质：

设 $X_1,\mathbf{L},X_r$ 之间相互独立，则若将它们分割成互不相交的 m 个块，则这 m 个块各自的函数 $Y_1,Y_2,\mathbf{L},Y_m$ 之间相互独立。

例 3.11

(1) **续例 3.5.** 有放回情形下 X_1，X_2 相互独立，无放回情形下不独立。

一般地，二个离散型随机变量相互独立等价于它们的联合分布列的各行成比例，或者各列成比例。

(2) 设 $(X_1, X_2) \sim U(D)$。当 D 是 $\mathbb{R}^2$ 中的矩形时，X_1，X_2 相互独立；反之，不独立。

(3) 设 $(\underset{\sim}{X}, \underset{\sim}{Y}) \sim N_k(m, \Sigma)$，其中 $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}$。则 $\underset{\sim}{X}$，$\underset{\sim}{Y}$ 相互独立等价于 $\Sigma_{12} = 0$。

§3.5 随机变量、向量函数的分布

在概率统计的理论与应用中，经常涉及到计算随机变量或随机向量的函数的分布问题。设 X 是定义在概率空间 (S, F, P) 上的一个 k 维随机向量，$g(\cdot): \mathbb{R}^k \to \mathbb{R}^m$ 是一个 m 维 $(m \le k)$ 的函数。记 $Y = g(X)$，那么，Y 是一个 m 维的随机向量，它的概率分布完全由 X 的概率分布决定。本节讨论如何根据 X 的分布来求 Y 的分布。

支撑（support）：设 $p(x)$ 是一个 pdf 或分布列，则称 $\mathrm{X} = \{x: p(x) > 0\}$ 为该分布的支撑。

若 $Y = g(X)$，X 是 X 分布的支撑。则 Y 分布的支撑为 $\mathrm{Y} = \{y: y = g(x),\ x \in \mathrm{X}\}$。

一、X 是离散型分布的情况

设 X 是离散型的 r. vec.，其分布列为 $P(X = x),\ x \in \mathrm{X}$。则 Y 也是一个离散型 r. vec。其分布列的计算步骤为：

1) 计算出 Y 的支撑 $\mathrm{Y} = \{y = g(x): x \in \mathrm{X}\}$；
2) 计算 $P(Y = y) = \sum_{\{x \in \mathrm{X}: g(x) = y\}} P(X = x),\quad \forall\, y \in \mathrm{Y}.$（把相同的 $g(x)$ 合并，对应的概率值相加。）

例 3.12 **(1)** 设 r. v. X 的分布列为

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

求 $Y = X^2 + X$ 的分布列。

解： 列表计算 Y 的可能取值：

X	-2	-1	0	1	2
$Y = X^2 + X$	2	0	0	2	6
P	0.2	0.1	0.1	0.3	0.3

得 Y 的分布列为

Y	0	2	6
P	0.2	0.5	0.3

(2) (Poisson 分布的独立可加性) 设 $X_j \sim \text{Poi}(l_j),\ j = 1,2$，且 X_1, X_2 相互独立，则 $Y = X_1 + X_2 \sim \text{Poi}(l_1 + l_2)$。

解： 显然，Y 的支撑为非负整数。

$$\begin{aligned} P(Y=k) &= P\left(\sum_{i=0}^{k}\{X_1 = i,\ X_2 = k-i\}\right) \\ &= \sum_{i=0}^{k} P\{X_1 = i,\ X_2 = k-i\} \quad \text{(卷积公式)} \\ &= \sum_{i=0}^{k} P(X_1 = i)P(X_2 = k-i) \ \text{(独立性)} \\ &= \text{.....} \\ &= \frac{1}{k!}(l_1 + l_2)^k e^{-(l_1+l_2)},\ k = 0,1,2,\mathbf{L} \end{aligned}$$

所以 $Y \sim \text{Poi}(l_1 + l_2)$。

一般地，若 $X_j \sim \text{Poi}(l_j),\ j = 1,2,\mathbf{L},n$，且 $X_1, X_2, \mathbf{L}, X_n$ 相互独立，则 $Y = \sum_{j=1}^{n} X_j \sim \text{Poi}(\sum_{j=1}^{n} l_j)$。

(3) **（二项分布的独立可加性）**若 $X_j \sim \text{B}(n_j, p),\ j = 1,2,\mathbf{L},m$，且 $X_1, X_2, \mathbf{L}, X_m$ 相互独立，则 $Y = \sum_{j=1}^{m} X_j \sim \text{B}(\sum_{j=1}^{m} n_j, p)$。

二、X 是连续型分布的情况

1. 一一对应变换

定理 3.3 设 r. vec. $X=(X_1,\mathbf{L}\,X_k)$ 具有 jpdf $p_X(x_1,\mathbf{L},x_k)$，其支撑为 $\mathbf{X}$。

$\begin{cases} y_1=g_1(x_1,\mathbf{L},x_k) \\ y_2=g_2(x_1,\mathbf{L},x_k) \\ \quad\mathbf{L}\quad\mathbf{L} \\ y_k=g_k(x_1,\mathbf{L},x_k) \end{cases}$ 在 $\mathbf{X}\to\mathbf{Y}$ 上一一对应，逆变换为 $\begin{cases} x_1=h_1(y_1,\mathbf{L},y_k) \\ x_2=h_2(y_1,\mathbf{L},y_k) \\ \quad\mathbf{L}\quad\mathbf{L} \\ x_k=h_k(y_1,\mathbf{L},y_k) \end{cases}$，

变换的 Jacobi 行列式存在，记为

$$J=\frac{\partial(x_1,\mathbf{L},x_k)}{\partial(y_1,\mathbf{L},y_k)}=\begin{vmatrix} \partial x_1/\partial y_1 & \partial x_2/\partial y_1 & \mathbf{L} & \partial x_k/\partial y_1 \\ \partial x_1/\partial y_2 & \partial x_2/\partial y_2 & \mathbf{L} & \partial x_k/\partial y_2 \\ \mathbf{L} & \mathbf{L} & \mathbf{L} & \mathbf{L} \\ \partial x_1/\partial y_k & \partial x_2/\partial y_k & \mathbf{L} & \partial x_k/\partial y_k \end{vmatrix}。$$

则随机向量 $\begin{cases} Y_1=g_1(X_1,\mathbf{L},X_k) \\ Y_2=g_2(X_1,\mathbf{L},X_k) \\ \quad\mathbf{L}\quad\mathbf{L} \\ Y_k=g_k(X_1,\mathbf{L},X_k) \end{cases}$ 的 jpdf 为

$$p_Y(y_1,\mathbf{L},y_k)=\begin{cases} p_X(h_1(y_1,\mathbf{L},y_k),\mathbf{L},h_k(y_1,\mathbf{L},y_k))\cdot|J|, & (y_1,\mathbf{L},y_k)\in\mathbf{Y}, \\ 0, & \text{else.} \end{cases}$$

例 3.13

(1) （**对数正态分布** LN(m, s^2)）设 $X\sim N(m,s^2)$, $Y=e^X$，求 Y 的 pdf。

解： 显然，$\mathbf{X}=\mathbb{R}$，$\mathbf{Y}=\mathbb{R}^+$，且 $y=e^x$ 是 $\mathbf{X}\to\mathbf{Y}$ 上一一对应的变换，逆变换为 $x=\log y$，$J=\dfrac{dx}{dy}=\dfrac{1}{y}$。因为 X 的 pdf 为

$$p_X(x)=\frac{1}{\sqrt{2p}\,s}e^{-\frac{(x-m)^2}{2s^2}},\quad x\in\mathbb{R},$$

故由定理 3.3 得 Y 的 pdf 为

$$p_Y(y)=\begin{cases} \dfrac{1}{\sqrt{2p}\,s}e^{-\frac{(\log y-m)^2}{2s^2}}\dfrac{1}{y}, & y\in\mathbb{R}^+, \\ 0, & \text{else.} \end{cases}$$

(2) （**逆 Gamma 分布**）设 $X \sim Gamma(a, b)$, $Y = \frac{1}{X}$，求 Y 的 pdf。

(3) （**多元正态的线性变换**）设 $X \sim N_k(m, \Sigma)$, $Y = AX + b$，其中 A 是给定的 k 阶非退化的方阵，b 是给定的 k 维列向量。则 $Y \sim N_k(Am + b, A\Sigma A')$。

2. $m<k$ **的变换**

方法一，补一些变换，构造成一一对应变换，然后通过边际分布求得。

2.1 求随机变量和的分布。 设 $(X_1, X_2) \sim p_X(x_1, x_2)$，求 $Y = X_1 + X_2$ 的分布。

先求 $\begin{cases} Y_1 = X_1 \\ Y = X_1 + X_2 \end{cases}$ 的 jpdf，然后对之积分获得 Y 的 pdf，即 (Y_1, Y) 的 mpdf。

由定理 3.3 得，

$$\begin{aligned} p_{Y_1,Y}(y_1, y) &= p_X(x_1(y_1, y), x_2(y_1, y))|J| \\ &= p_X(y_1, y - y_1), \end{aligned}$$

因此，Y 的 pdf 为

$$p_Y(y) = \int_{-\infty}^{\infty} p_X(y_1, y - y_1)dy_1 . \quad \text{（卷积公式）}$$

若 X_1, X_2 相互独立，则

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(y_1)p_{X_2}(y - y_1)dy_1 。$$

多个随机变量之和的分布也类似可求。

例 3.14

(1) （**多元正态的线性变换**）设 $X \sim N_k(m, \Sigma)$, $Y = AX + b$，其中 A 是给定的 $m \times k$ 阶行满秩矩阵，b 是给定的 m 维列向量，$m<k$。则 $Y \sim N_m(Am + b, A\Sigma A')$。

(2)（**独立 Gamma 分布之和**）设 $X_i \sim Gamma(a_i, b), i = 1, \mathbf{L}, k$，且 $X_1, \mathbf{L}, X_k$ 相互独立，则 $Y = \sum_{i=1}^{k} X_i \sim Gamma(\sum_{i=1}^{k} a_i, b)$。

2.2 求随机变量商的分布。设 $(X_1, X_2) \sim p_X(x_1, x_2)$，求 $U = X_1 / X_2$ 的分布。

先求 $\begin{cases} U = X_1 / X_2 \\ V = X_2 \end{cases}$ 的 jpdf，然后对之积分获得 U 的 pdf，即 (U,V) 的 mpdf。

由定理 3.3 得，

$$\begin{aligned} p_{U,V}(u,v) &= p_X(x_1(u,v), x_2(u,v))\,|J| \\ &= p_X(uv,v)\,|v|, \end{aligned}$$

因此，U 的 pdf 为

$$p_U(u) = \int_{-\infty}^{\infty} p_X(uv,v)\,|v|\,dv.$$

方法二，先求 U 的分布函数，再求 pdf。

U 的分布函数为

$$\begin{aligned} F_U(u) &= \iint_{x/y \le u} p_X(x,y)\,dxdy \\ &= \int_{-\infty}^{0}\left[\int_{uy}^{\infty} p_X(x,y)\,dx\right]dy + \int_{0}^{\infty}\left[\int_{-\infty}^{uy} p_X(x,y)\,dx\right]dy \\ &\overset{let\ t=x/y}{=} \int_{-\infty}^{0}\left[\int_{-\infty}^{u} p_X(ty,y)\,|y|\,dt\right]dy + \int_{0}^{\infty}\left[\int_{-\infty}^{u} p_X(ty,y)\,y\,dt\right]dy \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{u} p_X(ty,y)\,|y|\,dt\right]dy \\ &= \int_{-\infty}^{u}\left[\int_{-\infty}^{\infty} p_X(ty,y)\,|y|\,dy\right]dt, \quad \forall u \in \mathbb{R}. \end{aligned}$$

所以 U 的密度函数为

$$p_U(u) = \int_{-\infty}^{\infty} p_X(uy,y)\,|y|\,dy.$$

2.3 求随机变量最大值、最小值的分布。 设 $X_1, X_2, \mathbf{L}, X_n$ i.i.d. $f(x)$。记 $Y = \min(X_1, X_2, \mathbf{L}, X_n)$，$Z = \max(X_1, X_2, \mathbf{L}, X_n)$。求 Y，Z 的概率密度。

记 $f(x)$ 对应的分布函数为 $F(x)$。Y 的分布函数为

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\min\{X_1, \mathbf{L}, X_n\} \le y) \\ &= 1 - P(\min\{X_1, \mathbf{L}\ X_n\} > y) \\ &= 1 - P(\bigcap_{i=1}^{n}\{X_i > y\}) \\ &= 1 - \prod_{i=1}^{n} P(X_i > y) \\ &= 1 - [1 - F(y)]^n, \quad \forall y \in \mathbb{R}, \end{aligned}$$

对其求导可得 Y 的 pdf 为

$$f_Y(y)=n\left[1-F(y)\right]^{n-1}f(y),\ y\in \text{¡}.$$

Z 的分布函数为

$$\begin{aligned}F_Z(z)&=P(Z\le z)=P(\max\{X_1,\mathbf{L},X_n\}\le z)\\&=P(\prod_{i=1}^{n}\{X_i\le z\})\\&=\left[F(z)\right]^n,\quad \forall z\in \text{¡},\end{aligned}$$

对其求导可得 Z 的 pdf 为

$$f_Z(z)=n\left[F(z)\right]^{n-1}f(z),\ z\in \text{¡}.$$

2.4 求次序统计量的分布——概率元法

3. “多对一”的变换

定理 3.4 设 r. vec. $X=(X_1,\mathbf{L}\,X_k)$ 具有 jpdf $p_X(x_1,\mathbf{L},x_k)$，支撑为X。$\mathrm{X}_1,\mathbf{L},\mathrm{X}_r$ 是$\mathbf{X}$ 的一个分割。函数

$$\begin{cases}y_1=g_1(x_1,\mathbf{L},x_k)\\y_2=g_2(x_1,\mathbf{L},x_k)\\\quad\mathbf{L}\ \ \mathbf{L}\\y_k=g_k(x_1,\mathbf{L},x_k)\end{cases}$$ 在$\mathrm{X}_i\to\mathrm{Y}_i$ 上一一对应，逆变换为 $$\begin{cases}x_1=h_{i1}(y_1,\mathbf{L},y_k)\\x_2=h_{i2}(y_1,\mathbf{L},y_k)\\\quad\mathbf{L}\ \ \mathbf{L}\\x_k=h_{ik}(y_1,\mathbf{L},y_k)\end{cases},$$

变换的 Jacobi 行列式存在，记为 J_i，$i=1,\mathbf{L},r$。

则随机向量 $\begin{cases}Y_1=g_1(X_1,\mathbf{L},X_k)\\Y_2=g_2(X_1,\mathbf{L},X_k)\\\quad\mathbf{L}\ \ \mathbf{L}\\Y_k=g_k(X_1,\mathbf{L},X_k)\end{cases}$ 的 jpdf 为

$$p_Y(y_1,\mathbf{L},y_k)=\sum_{i=1}^{r}p_X(h_{i1}(y_1,\mathbf{L},y_k),\mathbf{L},h_{ik}(y_1,\mathbf{L},y_k))\cdot|J_i|I_{\mathrm{Y}_i}(y_1,\mathbf{L},y_k)。$$

第四章　数字特征

虽然，随机变量、随机向量的统计规律可由其概率分布全面地描述，但实际应用中往往需要用少量数值去概括一个概率分布的主要特征。这些数值就称为概率分布的数字特征。数字特征完全由分布决定

常用的数字特征有：

- 矩 (moments)，包括原点矩与中心矩两类，数学期望、方差、协方差是其中最常用的数字特征。
- 与矩有关的数字特征，如标准差、偏度系数、峰度系数、相关系数等。
- 分位点（quantiles），常用的有第一四分位数、第三四分位数、中位数等。
- 与分位点有关的数字特征，如四分位间距、极差等。
- 众数。

§4.1　数学期望 (Expectation)

数学期望是最重要、最基本的数字特征。直观含义是将随机变量的可能取值按其相应的概率加权平均。是刻划随机变量分布的中心位置的一种办法。源于 17 世纪 Pascal 对于分赌注问题的研究。

（**分赌注问题**）甲乙两人各出赌注 50 法郎，用轮流掷一枚均匀硬币进行赌博。规则是掷一次硬币算一局，得正面算甲胜，得反面算乙胜。谁先胜三局就赢得全部 100 法郎赌注。掷三次硬币后，甲胜两局乙胜一局，赌博因故中止。问这 100 法郎赌注该如何分配才算合理？

定义 4.1 设 r.v. X 的分布函数为 $F(x)$，$g(x):\mathbb{R}\to\mathbb{R}$ 是一已知函数，若 Lebesgue－Stieltjes 积分 $\int_{-\infty}^{\infty}|g(x)|dF(x)<\infty$，则称 L-S 积分 $\int_{-\infty}^{\infty}g(x)dF(x)$ 为 $g(X)$ 的数学期望，记为 $Eg(X)$。

注 4.1：

(1) $g(x)\equiv x$ 时，即得 EX 。

(2) L-S 积分是 Riemann-Stieltjes 积分的推广。$[a,b]$ 上 R-S 积分 $\int_a^b g(x)dF(x)$ 的含义是：和式 $\sum_{i=1}^{n}g(x_i)[F(x_i)-F(x_{i-1})]$ 在分割细度趋于零时的极限。

(3) 当 $F(x)$ 为阶梯函数时，即 X 具有离散型分布，设其分布列为 $\{P(X=x): x\in \mathrm{X}\}$，则 $Eg(X)=\sum_{x\in \mathrm{X}} g(x)P(X=x)$。

(4) 若 X 具有连续型分布，$p(x)$ 为其密度函数，则 $Eg(X)=\int_{-\infty}^{\infty} g(x)p(x)dx$，此处的积分往往等同于 Riemann 积分。

(5) 若记 $Y=g(X)$，其分布函数为 $F_Y(y)$，则可证明

$$EY=\int_{-\infty}^{\infty} ydF_Y(y)=\int_{-\infty}^{\infty} g(x)dF(x)。$$

定义 4.2 设 r.vec. $(X_1,\mathbf{L},X_n)$ 的分布函数为 $F(x_1,\mathbf{L},x_n)$，$g(x_1,\mathbf{L},x_n): \mathbb{R}^n\to\mathbb{R}$ 是一已知函数，若 Lebesgue－Stieltjes 积分 $\int_{\mathbb{R}^n} |g(x_1,\mathbf{L},x_n)|dF(x_1,\mathbf{L},x_n)<\infty$，则称 L-S 积分 $\int_{\mathbb{R}^n} g(x_1,\mathbf{L},x_n)dF(x_1,\mathbf{L},x_n)$ 为 $g(X_1,\mathbf{L},X_n)$ 的数学期望，记为 $Eg(X_1,\mathbf{L},X_n)$。称 $(EX_1,\mathbf{L},EX_n)$ 为 r.vec. $(X_1,\mathbf{L},X_n)$ 的期望向量。

注 4.2：

(1) 若记 $Y=g(X_1,\mathbf{L},X_n)$，其分布函数为 $F_Y(y)$，则可证明

$$EY=\int_{-\infty}^{\infty} ydF_Y(y)=\int_{\mathbb{R}^n} g(x_1,\mathbf{L},x_n)dF(x)。$$

特别地，取 $g(x_1,\mathbf{L},x_n)=x_j$，则得

$$EX_j=\int_{\mathbb{R}^n} x_j dF(x_1,\mathbf{L},x_n)=\int_{-\infty}^{\infty} x_j dF_j(x_j)。$$

(2) 当 $(X_1,\mathbf{L},X_n)$ 具有离散型分布时，

$$Eg(X_1,\mathbf{L},X_n)=\sum_{(x_1,\mathbf{L},x_n)\in \mathrm{X}} g(x_1,\mathbf{L},x_n)P(X_1=x_1,\mathbf{L},X_n=x_n)。$$

(3) 当 $(X_1,\mathbf{L},X_n)$ 具有连续型分布，$p(x_1,\mathbf{L},x_n)$ 为其密度函数，则

$$Eg(X_1,\mathbf{L},X_n)=\int_{-\infty}^{\infty}\mathbf{L}\int_{-\infty}^{\infty} g(x_1,\mathbf{L},x_n)p(x_1,\mathbf{L},x_n)dx_1\,\mathbf{L}\,dx_n,$$

此处的积分一般等于 n-重积分。

数学期望的性质

(1) 若 $a \le g(x) \le b, \forall x$，则 $a \le Eg(X) \le b$。特别地，$Ec = c$。

(2) （线性性质）$E\left[ag_1(X) + bg_2(X) + c\right] = aEg_1(X) + bEg_2(X) + c$。

(3) 若 $g_1(x) \ge g_2(x), \forall x$，则 $Eg_1(X) \ge Eg_2(X)$。

(4) 若 $\underset{\sim}{X}_1, \mathbf{L}, \underset{\sim}{X}_r$ 相互独立，$\underset{\sim}{X}_j$ 的维数为 k_j，$f_j(\cdot): \mathbb{R}^{k_j} \to \mathbb{R}$，$j = 1, \mathbf{L}, r$，则

$$E\left[f_1(\underset{\sim}{X}_1) \mathbf{L}\ f_r(\underset{\sim}{X}_r)\right] = Ef_1(\underset{\sim}{X}_1) \mathbf{L}\ Ef_r(\underset{\sim}{X}_r)。$$

注 4.3 性质中 X 可以是 n-维随机向量，相应的 $g(\cdot),\ g_1(\cdot),\ g_2(\cdot)$ 为 $\mathbb{R}^n \to \mathbb{R}$ 的函数。

例 4.1

(1) 设 $X \sim B(n, p)$，则 $EX = np$。

(2) 设 $X \sim Poi(l)$，则 $EX = l$。

(3) 设 $X \sim Ge(p)$，则 $EX = \dfrac{1}{p}$。

(4) 设 $X \sim U(a, b)$，则 $EX = \dfrac{a+b}{2}$。

(5) 设 $X \sim Exp(l)$，则 $EX = \dfrac{1}{l}$。

(6) 设 $X \sim N(m, s^2)$，则 $EX = m$.

(7) 设 X 的分布列为 $P\left[X = (-1)^k 2^k / k \right] = 2^{-k},\ k = 1, 2, \mathbf{L}$。虽然

$$\sum_{k=1}^{\infty} (-1)^k 2^k / k \cdot 2^{-k} = -\log 2,$$

但该级数不绝对收敛，所以 X 的期望不存在。

(8) 设 X 服从 Cauchy 分布，pdf 为 $p(x) = \dfrac{1}{p(1+x^2)}, x \in \mathbb{R}$，则 E$X$ 不存在。

(9) 设 $X \sim N(m, s^2),\ Y = e^X \sim LN(m, s^2)$，则

$$EY = Ee^X = \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2p}\, s} e^{-\frac{1}{2s^2}(x-m)^2} dx = e^{m + \frac{s^2}{2}}。$$

(10) 设 X_1, X_2 *i.i.d.* $\mathrm{Exp}(l)$, $Y=\max(X_1,X_2)$，则

$$\begin{aligned} EY &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\max(x_1,x_2)\,p(x_1)\,p(x_2)dx_1dx_2 \\ &= \int_0^{\infty}\int_0^{x_1} x_1 p(x_1)\,p(x_2)dx_1dx_2 + \int_0^{\infty}\int_{x_1}^{\infty} x_2 p(x_1)\,p(x_2)dx_1dx_2 \\ &= \frac{3}{2l}. \end{aligned}$$

(11) 设 $X \sim HG(n;N,M)$，即分布列为

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=\max\{0,n-(N-M)\},\mathbf{L},\min\{M,n\}$$

则 $EX=n\cdot\frac{M}{N}$。

(12) $X\sim c^2(n),\ EX=n;\quad X\sim t(n),\ EX=0\ (n>1);\quad X\sim F(m,n),\ EX=\frac{n}{n-2}\ (n>2).$

§4.2 方差 (Variance)

定义 4.3 若 $E(X-EX)^2$ 存在，则称之为 r.v. X 的方差，记为 $\mathrm{Var}(X)$。称 $\sqrt{\mathrm{Var}(X)}$ 为 X 的标准差（standard deviation, std）。

方差是刻划随机变量取值分散程度的一个量，方差越小说明 X 的取值越集中，越大则越分散。方差存在则期望一定存在；反之不然。

计算方法

- $\mathrm{Var}(X)=\begin{cases}\sum\limits_{x\in X}(x-EX)^2P(X=x), & X\text{离散型分布},\\ \int_{-\infty}^{\infty}(x-EX)^2\,p(x)dx, & X\text{连续型分布}.\end{cases}$
- $\mathrm{Var}(X)=EX^2-(EX)^2$

方差的性质

(1) c 为常数，则 $\mathrm{Var}(c)=0$。

(2) a，c 为常数，则 $\mathrm{Var}(aX+c)=a^2\,\mathrm{Var}(X)$。

(3) $\mathrm{Var}(X)=\min\limits_{c} E(X-c)^2$。

(4) 若 $X_1,\mathbf{L},X_n$ 相互独立，则 $\mathrm{Var}(\sum\limits_{i=1}^{n}X_i)=\sum\limits_{i=1}^{n}\mathrm{Var}(X_i)$。

(5) (Chebyshev 不等式) 若 $m=EX$，$s^2=\mathrm{Var}(X)$，则对 $\forall t>0$，有

$$P\left(|X-m|\geq t\right)\leq\frac{s^2}{t^2}。$$

若令 $t=cs$，则有

$$P\left(|X-m|\geq cs\right)\leq\frac{1}{c^2}。$$

Chebyshev 不等式的一般形式：

设 X 是一个 r.v.，$g(x)$是一个非负函数，则对 $\forall r>0$，有 $P(g(X)\geq r)\leq\dfrac{Eg(X)}{r}$。

(6) $Var(X)=0\iff P(X=EX)=1.$

随机变量的标准化

若 $m=EX$，$s^2=\mathrm{Var}(X)>0$，则称 $X^*=\dfrac{X-m}{s}$ 是 X 的标准化。易知，$EX^*=0,\ \mathrm{Var}(X^*)=1$。

例 4.2

(1) 设 $X\sim B(n,p)$，则 $EX=np$，$Var(X)=np(1-p)$。

(2) 设 $X\sim Poi(l)$，则 $EX=l$，$Var(X)=l$。

(3) 设 $X\sim Ge(p)$，则 $EX=\dfrac{1}{p}$，$Var(X)=\dfrac{1-p}{p^2}$。

(4) 设 $X\sim U(a,b)$，则 $EX=\dfrac{a+b}{2}$，$Var(X)=\dfrac{1}{12}(b-a)^2$。

(5) 设 $X\sim Exp(l)$，则 $EX=\dfrac{1}{l}$，$Var(X)=\dfrac{1}{l^2}$。

(6) 设 $X \sim N(m, s^2)$，则 $EX = m$， $Var(X) = s^2$。

(7)

$X \sim c^2(n),\ EX = n,\ Var(X) = 2n;$

$X \sim t(n),\quad EX = 0\ (n > 1),\quad n > 2$时 $Var(X) = \dfrac{n}{n-2};$

$X \sim F(m,n),\ EX = ?\ \ Var(X) = ?$

§4.3 协方差与相关系数

协方差与相关系数是刻划两个 r.v. 之间线性相关程度的数字特征。

定义 4.4 若 $E(X - EX)(Y - EY)$ 存在，则称之为 r.v. X 与 Y 的协方差，记为 $Cov(X,Y)$。

定义 4.5 若 $Var(X) \cdot Var(Y) > 0$，则称 $\dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ 为 X 与 Y 的相关系数，记为 $r_{X,Y}$。

注：

- $Cov(X,Y)(or\ \ r_{X,Y})\begin{cases} > 0, & \text{称 } X,Y \text{ 正相关} \\ = 0, & \text{称 } X,Y \text{ 不相关} \\ < 0, & \text{称 } X,Y \text{ 负相关} \end{cases}$
- 设 X^*, Y^* 分别是 r.v. X, Y 的标准化，则 $Cov(X^*, Y^*) = r_{X,Y} = r_{X^*,Y^*}$。

性质

1. $Cov(X,Y) = Cov(Y,X)$。
2. $Cov(X,Y) = E(XY) - EX \cdot EY$。
3. $Cov(X,X) = Var(X);\ \ Cov(X,c) = 0, \forall$常数$c$。
4. $Cov(aX + b, cY + d) = acCov(X,Y)$。
5. $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$。
6. 与方差的关系：$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)$。
7. (Cauchy-Schwartz 不等式)

   对任意随机变量 X,Y 有 $\left[E(XY)\right]^2 \le EX^2 \cdot EY^2$；等号成立当且仅当存在不全为 0 的

$a, b,$ 使得 $P(aX+bY=0)=1$。

推论 1. $\left[Cov(X,Y)\right]^2 \le Var(X)\cdot Var(Y)$；等号成立当且仅当存在不全为 0 的 $a, b,$ 使得 $P(aX+bY=c)=1$。

推论 2. $\forall X,Y, \ |r_{X,Y}|\le 1;$

$$r_{X,Y}=\pm 1 \Longleftrightarrow \exists a\neq 0, b, \ \ni P(Y=aX+b)=1,$$

且当 $a>0$ 时 $r_{X,Y}=1,$ 当 $a<0$ 时 $r_{X,Y}=-1.$

8. (独立与不相关) 若 X, Y 独立，则 $Cov(X,Y)=0$；反之不成立。

例 4.3 设 (X,Y) 服从单位圆内的均匀分布，则 $Cov(X,Y)=0$ 但两者不独立。

例 4.4 设 (X,Y) 服从二元正态分布，则 X 与 Y 相互独立等价于 $Cov(X,Y)=0$。

定义 4.6 设 $\underset{\sim}{X}=(X_1,\mathbf{L},X_n)'$ 是一个随机向量，记 $b_{ij}=Cov(X_i,X_j)$，称 $(b_{ij})_{n\times n}\triangleq Cov(\underset{\sim}{X},\underset{\sim}{X})$ 为 $\underset{\sim}{X}$ 的协方差矩阵；记 $r_{ij}=r_{X_i,X_j}$，称 $(r_{ij})_{n\times n}\triangleq Corr(\underset{\sim}{X},\underset{\sim}{X})$ 为 $\underset{\sim}{X}$ 的相关系数矩阵。

性质：

若 $Cov(X_i,X_j), i,j=1,\mathbf{L},n$ 存在，则 $Cov(\underset{\sim}{X},\underset{\sim}{X}),\ Corr(\underset{\sim}{X},\underset{\sim}{X})$ 非负定。

例 4.5 若 $X\sim N_k(m,\Sigma)$，则

$EX=m;\ \ Cov(X,X)=\Sigma;$

$X_1,\mathbf{L},X_k$ 两两不相关等价于 $X_1,\mathbf{L},X_k$ 独立.

§4.4 其他数字特征

矩及与矩有关的数字特征

k 阶原点矩：$EX^k,\ k=1,2,\mathbf{L}$

k 阶中心矩：$E(X-EX)^k,\ k=1,2,\mathbf{L}$

偏度系数：$g_3=\dfrac{E(X-EX)^3}{\left[Var(X)\right]^{3/2}}$

峰度系数：$g_4=\dfrac{E(X-EX)^4}{\left[Var(X)\right]^2}-3$

变异系数：$CV=\dfrac{\sqrt{Var(X)}}{EX}\ (EX>0)$

对于随机向量，还常用到各阶混合矩。

条件期望与条件方差

$$E(X)=E\left[E(X\mid Y)\right]$$
$$Var(X)=E\left[Var(X\mid Y)\right]+Var\left[E(X\mid Y)\right]$$

与数学期望有关的常用不等式

1. Chebychev's Inequality

设 X 是一个 r.v.，g(x)是一个非负函数，则对 $\forall r>0$，有 $P(g(X)\ge r)\le\dfrac{Eg(X)}{r}$。

2. Hölder Inequality

设 X,Y 是任意两个随机变量，常数 p,q 满足 $\dfrac{1}{p}+\dfrac{1}{q}=1$，则

$$|EXY|\le E|XY|\le\left(E|X|^p\right)^{\frac{1}{p}}\left(E|Y|^q\right)^{\frac{1}{q}}。$$

3. Cauchy-Schwartz Inequality

设 X,Y 是任意两个随机变量，则

$$|EXY|\le E|XY|\le\left(E|X|^2\right)^{\frac{1}{2}}\left(E|Y|^2\right)^{\frac{1}{2}}。$$

4. Minkowski Inequality

设 X, Y 是任意两个随机变量，则对 $\forall\ p \in [1, \infty)$ 有

$$\left[E\,|\,X+Y\,|^{p}\right]^{\frac{1}{p}} \le \left(E\,|\,X\,|^{p}\right)^{\frac{1}{p}} + \left(E\,|\,Y\,|^{p}\right)^{\frac{1}{p}}。$$

5. Jensen Inequality

设 $g(x)$ 为 (a,b) 上的凸函数，X 是 (a,b) 上的随机变量且 EX 有限，则

$$g(EX) \le Eg(X)。$$

若 $g(x)$ 严凸且 $P(X \ne EX) > 0$，则不等号严格成立。

（多元情形） 随机向量 X 定义于开凸集 $S \in \mathbb{R}^{k}$ 上，且 X 的期望向量 EX 存在、有限，$g(x)$ 为 S 上的凸函数，则 $g(EX) \le Eg(X)$。若 $g(x)$ 严凸且 $P(X \ne EX) > 0$，则不等号严格成立。

第五章　特征函数

特征函数是分布函数的 Fourier 变换，是全面描述概率分布的又一个工具，在处理有些问题时非常方便。它能把求解独立同分布随机变量和的分布的卷积运算转换成乘法运算，能把求分布各阶矩的运算转换成微分运算，还能把求随机变量序列极限分布的问题转换为一般的函数极限问题。

一、随机变量的特征函数

定义 5.1　(特征函数) 设 X 是一个随机变量，分布函数为 $F(x)$ ，则称

$$j(t)=E\left(e^{itX}\right)=\int_{-\infty}^{\infty}e^{itx}dF(x),\quad t\in(-\infty,\infty)$$

为 X (或 $F(x)$) 的特征函数 (characteristic function)。其中 $i=\sqrt{-1}$ 是虚数单位。

注：

- 由欧拉公式知 $e^{itx}=\cos(tx)+i\sin(tx)$，因而 $j(t)=E\cos(tX)+iE\sin(tX)$，它是一个实变量的复值函数。由于 $|e^{itx}|=1$，所以特征函数 $j(t)$ 对 $\forall\, t\in(-\infty,\infty)$ 都有意义。
- 特征函数只依赖于分布函数，由分布所决定。
- 若 X 服从离散型分布，分布列为 $\{P(X=x_j),\ j=1,2,\mathbf{L}\}$，则 X 的特征函数为

$$j(t)=\sum_{j=1}^{\infty}e^{itx_j}P(X=x_j),\quad t\in\ ¡\ ;$$

若 X 具有连续型分布，密度函数为 $p(x)$，则 X 的特征函数为

$$j(t)=\int_{-\infty}^{\infty}e^{itx}p(x)dx,\quad t\in\ ¡\ 。$$

特征函数的性质

设 $j(t)$ 是某个随机变量的特征函数。

性质 1. $|j(t)|\le j(0)=1,\quad j(-t)=\overline{j(t)}$.

性质 2. $j(t)$ 在 $(-\infty,\infty)$ 上一致连续。

性质 3. **(非负定性)** 对任意正整数 n、任意实数 $t_1, \mathbf{L}, t_n$，及任意复数 $z_1, \mathbf{L}, z_n$，有

$$\sum_{u=1}^{n}\sum_{v=1}^{n} j(t_u - t_v) z_u \overline{z}_v \geq 0 \text{。}$$

性质 4. 设 $Y = aX + b$，其中 a, b 为常数，则 Y 的特征函数为 $j_Y(t) = e^{ibt} j_X(at)$。

性质 5. 两个独立随机变量之和的特征函数等于它们的特征函数之积，即若 X 与 Y 相互独立，则 $j_{X+Y}(t) = j_X(t) \cdot j_Y(t)$。该性质亦可推广到 n 个独立随机变量之和的情形。

性质 6. 若 X 的 k 阶矩存在，则 X 的特征函数 $j(t)$ 可 k 次微分，对 $m \leq k$ 有

$$j^{(m)}(0) = i^m E(X^m),$$

且有如下展式

$$j(t) = 1 + (it)EX + \frac{(it)^2}{2!}EX^2 + \mathbf{L} + \frac{(it)^k}{k!}EX^k + o(t^k) \text{。}$$

逆转公式与唯一性定理

定理 5.1 **(逆转公式)** 设随机变量 X 的分布函数为 $F(x)$，特征函数为 $j(x)$，则对 $F(x)$ 的任意两个连续点 $x_1 < x_2$ 有

$$F(x_2) - F(x_1) = \lim_{T \to \infty} \frac{1}{2p} \int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{it} j(t) dt \text{。}$$

定理 5.2 **(唯一性定理)** 随机变量的分布函数由其特征函数唯一决定。

定理 5.3 若特征函数 $j(t)$ 绝对可积，即 $\int_{-\infty}^{\infty} |j(t)| dt < \infty$，则其相应的分布函数 $F(x)$ 的导数存在且连续，而且

$$F'(x) = \frac{1}{2p} \int_{-\infty}^{\infty} e^{-itx} j(t) dt \text{。}$$

常用分布的特征函数

(1) 退化分布 $P(X = c) = 1$：$j(t) = e^{ict}$。

(2) $B(n,p)$： $j(t)=\left[pe^{it}+(1-p)\right]^n$。

(3) $Poi(l)$： $j(t)=e^{l(e^{it}-1)}$。

(4) $U(a,b)$： $j(t)=\dfrac{e^{ibt}-e^{iat}}{it(b-a)}$。

(5) $N(0,1)$: $j(t)=e^{-t^2/2}$； $N(m,s^2)$: $j(t)=e^{imt-\frac{s^2t^2}{2}}$。

(6) $Exp(l)$: $j(t)=(1-\dfrac{it}{l})^{-1}$。

(7) $Gamma(a,b)$: $j(t)=\left(1-itb\right)^{-a}$。

(8) $t(1)$ (Cauchy distribution)**:** $j(t)=e^{-|t|}$。

二、随机向量的特征函数

定义 5.2 (随机向量的特征函数) 设随机向量 $(X_1,\mathbf{L},X_k)$ 的分布函数为 $F(x_1,\mathbf{L},x_k)$，则称

$$j(t_1,\mathbf{L},t_k)=E\left[e^{i(t_1X_1+\mathbf{L}+t_kX_k)}\right]=\int_{-\infty}^{\infty}\mathbf{L}\int_{-\infty}^{\infty}e^{i(t_1x_1+\mathbf{L}+t_kx_k)}dF(x_1,\mathbf{L},x_k),\quad (t_1,\mathbf{L},t_k)\in\mathbf{¡}^k$$

为 $(X_1,\mathbf{L},X_k)$ (或$F(x_1,\mathbf{L},x_k)$) 的特征函数。

性质

设 $j(\underset{\sim}{t}),\ \underset{\sim}{t}\in\mathbf{¡}^k$ 是 k 维随机向量的特征函数。

(1) $|j(\underset{\sim}{t})|\le j(\underset{\sim}{0})=1,\quad j(-\underset{\sim}{t})=\overline{j(\underset{\sim}{t})}$.

(2) $j(\underset{\sim}{t})$ 在 $\mathbf{¡}^k$ 上一致连续。

(3) 两个相互独立、维数相同的随机向量之和的特征函数等于它们的特征函数之积，即若 X 与 Y 相互独立，则 $j_{X+Y}(\underset{\sim}{t})=j_X(\underset{\sim}{t})\cdot j_Y(\underset{\sim}{t})$。该性质可推广到 n 个独立随机向量之和的情形。

(4) **（唯一性定理）**随机向量的分布函数与特征函数一一对应。

(5) 若矩 $E\left(X_1^{m_1}\,\mathbf{L}\,X_k^{m_k}\right)$ 存在，则

$$E\left(X_1^{m_1}\,\mathbf{L}\,X_k^{m_k}\right)=i^{-(m_1+\mathbf{L}+m_k)}\left[\frac{\partial^{m_1+\mathbf{L}+m_k}\,j\,(\underline{t})}{\partial t_1^{m_1}\,\mathbf{L}\,\partial t_k^{m_k}}\right]_{\underline{t}=\underline{0}}。$$

(6) 设 k 维随机向量 $X=(X^{(1)},X^{(2)})\sim j\,(\underline{t}^{(1)},\underline{t}^{(2)})$，其中 m（$m<k$）维子向量 $X^{(1)}\sim j_1(\underline{t}^{(1)})$，$k-m$ 维子向量 $X^{(2)}\sim j_2(\underline{t}^{(2)})$。则

a) $X^{(1)}$ 的特征函数 $j_1(\underline{t}^{(1)})=j\,(\underline{t}^{(1)},\underline{0})$；

b) $X^{(1)},X^{(2)}$ 相互独立当且仅当 $j\,(\underline{t}^{(1)},\underline{t}^{(2)})=j_1(\underline{t}^{(1)})j_2(\underline{t}^{(2)})$。

(7) 若 $X\sim j\,(\underline{t}),\ Y=AX+\underline{b}$，则 Y 的特征函数为 $j_Y(\underline{t})=e^{i\underline{b}'\underline{t}}j_X(A'\underline{t})$。

(8) k 维随机向量 X 的分布由一切形如 $\underline{a}'X,\ \underline{a}\in\mathbf{¡}^k$ 的分布唯一决定。

证：因为 $\underline{a}'X$ 的特征函数为 $j_{\underline{a}}(t)=E\left[e^{it(\underline{a}'X)}\right]$。取 $t=1$ 得 $j_{\underline{a}}(1)=E\left[e^{i\underline{a}'X}\right]$，把它看作 $\underline{a}$ 的函数正好是 X 的特征函数。由唯一性定理知，它决定了 X 的分布。

多元正态分布的特征函数

$X\sim N_k(\underline{m},\Sigma)$，则 X 的特征函数为 $j\,(\underline{t})=e^{i\underline{t}'\underline{m}-\frac{1}{2}\underline{t}'\Sigma\underline{t}}$。

三、其他生成函数

矩母函数 (moment generating function)：设 $X\sim F(x)$，若存在 $h>0$，使得 $\forall t\in(-h,h)$，Ee^{tX} 存在，则称 $M_X(t)=Ee^{tX}$ 为 X (或 $F(x)$) 的矩母函数。

累积量生成函数 (cumulant generating function)：$K(t)=\log[M_X(t)]$。

第六章　概率极限定理

一、　依概率收敛 (Convergence in Probability)

定义 6.1 （**依概率收敛**）设 $X_1, X_2, \mathbf{L}$ 是一个随机变量序列，X 是一个随机变量。若对 $\forall e > 0$，有

$$\lim_{n\to\infty} P(|X_n - X| < e) = 1,$$

则称序列 $\{X_n\}$ 依概率收敛于 X，记作 $X_n \to_P X$ 。

注：

序列 $\{X_n\}$ 服从大数定律(Law of Large Numbers) 等价于 $Y_n \to_P 0$，其中 $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i - a_n$。

常见的大数定律

1. (常用的形式) 设 $X_1, X_2, \mathbf{L}$ i.i.d.，$EX_1 = m$, $Var(X_1) = s^2 < \infty$，记 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$。则 $\bar{X}_n \to_p m$。（用 Chebyshev 不等式可证）

2. (Chebyshev LLN) 设 $X_1, X_2, \mathbf{L}$ 是一列两两不相关的随机变量序列，各个变量的方差存在且有共同上界，即 $Var(X_i) \le c,\ i = 1, 2, \mathbf{L}$。则

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \to_p 0。$$

3. （**Markov LLN**）设 $X_1, X_2, \mathbf{L}$ 是一列两两不相关的随机变量序列，各个变量的方差存在，且满足 $\frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \to 0$。则 $\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \to_p 0$。

4.（**辛钦 LLN**）设 $X_1, X_2, \mathbf{L}$ i.i.d.，若 $EX_1 = m$ 存在，则 $\bar{X}_n \to_p m$。

定理 6.1 （斯鲁茨基定理）设 $\{X_{1,n}, n\geq 1\}, \{X_{2,n}, n\geq 1\}, \mathbf{L}, \{X_{k,n}, n\geq 1\}$ 是 k 个随机变量序列，且

$$X_{j,n} \to_P a_j \ (n\to\infty), \quad j=1,\mathbf{L},k,$$

又 $R(x_1,\mathbf{L},x_k)$ 是 k 元有限的有理函数，则有

$$R(X_{1,n},\mathbf{L},X_{k,n}) \to_P R(a_1,\mathbf{L},a_k)。$$

定理 6.2 设 $X_n \to_P X$ ， $h(\cdot)$ 是一个连续函数，则 $h(X_n) \to_P h(X)$ 。

二、 几乎处处收敛 (Almost Sure Convergence)

定义 6.2 （几乎处处收敛）设 $X_1, X_2, \mathbf{L}$ 是一个随机变量序列，X 是一个随机变量。若对 $\forall e>0$，有

$$P(\lim_{n\to\infty} |X_n - X| < e) = 1,$$

则称序列 $\{X_n\}$ 几乎处处收敛于 X，记作 $X_n \to X \ a.s.$ 。

注：

- 几乎处处收敛强于依概率收敛。通常前者可以推出后者，反之不行。
- 一个依概率收敛的序列，可以找到一个几乎处处收敛的子列。

定理 6.3 (强大数定律) 设 $X_1, X_2, \mathbf{L}$ i.i.d.， $EX_1 = m$ 存在，记 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 。则 $\bar{X}_n \to m \ \ a.s.$ 。

三、 分布收敛 (Convergence in Distribution)

定义 6.3 （分布收敛）设 $F(x), F_1(x), F_2(x), \mathbf{L}$ 是一个分布函数的序列，若对 $F(x)$ 的每个连续点 x 都有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则称分布函数序列 $\{F_n(x)\}$ 弱收敛于 $F(x)$，记作 $F_n(x) \to_W F(x)$ 。

若随机变量序列 $X_1, X_2, \mathbf{L}$ 的分布函数 $F_n(x)$ 弱收敛于随机变量 X 的分布函数，则称序列 $\{X_n\}$ 按分布收敛于 X，记作 $X_n \to_D X$。

定理 6.4 若 $X_n \to_P X$，则 $X_n \to_D X$。

定理 6.5 $X_n \to_P c$ (c 为常数) $\Leftrightarrow\Rightarrow F_n(x) \to_W F(x)$，其中 $F_n(x)$ 是 X_n 的分布函数，$F(x)$ 是退化分布 $P(X=c)=1$ 的分布函数。

定理 6.6 分布函数序列 $\{F_n(x)\}$ 弱收敛于分布函数 $F(x)$ 的充要条件是相应的特征函数序列 $\{j_n(t)\}$ 收敛于 $F(x)$ 的特征函数 $j(t)$。

中心极限定理的一个常用的形式。设 $X_1, X_2, \mathbf{L}$ i.i.d.，$EX_1 = m$，$0 < Var(X_1) = s^2 < \infty$，记 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$。则

$$\lim_{n\to\infty} P\{\sqrt{n}(\bar{X}_n - m)/s \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2p}} e^{-y^2/2} dy \text{。}$$

定理 6.7 (Slutsky) 若 $X_n \to_D X$，$Y_n \to_P a$，则

a) $Y_n X_n \to_D aX$；

b) $X_n + Y_n \to_D X + a$。

第七章　样本与统计量

一、　统计学概述

1. 统计学(statistics) 的由来

统计源于社会活动中的计数需要，历史可以追溯到原始社会。早期主要为统治者了解人口、土地、财富等国情的需要服务。

对 statistics 一词的由来说法不一，例如有

- 源于意大利，统计学即国情学（陈希孺《机会的数学》，p55）；
- 源于 1660 年德国人 Hermann Conring（Iversen, G. R.等著，吴喜之等译《统计学－－基本概念和方法》，p2）；
- 源于十八世纪中叶德国学者 G. Achenwall（C.R.Rao,《统计与真理－－怎样运用偶然性》，p31);

等等。

2. 统计学

统计学是一门关于收集和分析数据的科学和艺术。(《不列颠百科全书》)

统计学内容包括：收集数据、分析数据、并从中得到有用信息为决策提供依据的一系列概念、原则和方法。是一门研究数据的学问。

3. 数理统计(mathematical statistics)

以研究推断方法为主要目的，是统计学的一个分支，也是现代各类统计学的基础。

研究的数据 —— 带有随机性的数据。

研究的方法 —— 以概率论等数学方法为工具

研究的内容

- 如何有效地收集数据：采用各种随机抽样方法收集数据，使数据代表性好、尽量避免主观干扰。有抽样调查、试验设计两个分支。
- 分析数据：将研究对象的全体视为总体，将获得数据的部分个体视为样本，采用数学模型描述总体、样本及两者间的关系，由样本数据推断总体信息。
  分析方法有参数估计、假设检验、回归分析、时间序列分析、多元分析、非参数分析等等

与概率论的差别：是归纳推理而不是演绎。可视为概率论的一种应用。

本课程中，主要介绍一些统计分析的思想与方法，对数据收集方法不作详细讨论。

4. 统计学的应用领域

应用涉及：社会、经济状况调查，民意测验、市场调查、收视率调查，保险、金融，司法，交通，气象、地质，医药、疾病研究，生物、遗传研究，心理学、教育学研究，语言学研究，航天航空，质量管理等等。

与具体应用领域相结合，形成特定领域的统计学分支学科，如：计量经济、金融统计、教育统计、国民经济统计、可靠性统计、生存分析、保险统计、生物医药统计等等，这些分支学科大多以数理统计为基础。

二、 总体与样本

总体(population)

一项统计研究关心的往往是某个特定群体的整体信息。我们称该特定的群体为**总体**。

总体由**个体**(element)构成。

个体是承载数据的基本单位。数据收集是针对个体进行的。但是，统计研究中，收集个体的数据不是为了研究个体本身，而是为了从个体数据中得到总体的信息。

变量：一项统计研究通常不会对每个个体的一切方面都感兴趣，而只对它的某一项或几项数量指标感兴趣。由于一个数量指标在各个体上的取值往往是不同的，因而常称为**变量**，用大写英文字母 X, Y,... 等表示。

总体分布

欲研究的某些变量在总体中取各种值的比例分布称为**总体分布**（类似于概率分布）。统计研究主要关心的是某些变量的总体分布情况，所以常把总体分布与总体等同看待。

总体分布族

对欲研究的总体分布，往往有些信息已知，有些信息未知、需要推断。数理统计中，往往会根据一些已知的信息，对总体分布作一个合理的模型假定。即假定待研究的总体分布是某个分布族 $\{F(x;q): q \in \Theta\}$ 中的一员，其中 q 表示总体分布中的未知的信息。总体分布究竟是这个分布族中的哪一个，取决于 q 。如果q 是一个维数有限的未知参数向量，则称相应的总体分布族为*参数型*的；否则，称为*非参数型*的。

例 7.1 (1) 为了解上海市居民家庭经济情况，从市区常住户口中随机选取 500 户，对他们去年的收支情况进行详细记录，获得统计数据。根据这些数据对全市家庭去年的收支情况作推断。

总体：全部上海市的居民家庭。个体：家庭。变量：各项收入、支出。

(2) 收视率调查。AC 尼尔森公司从上海市居民中随机选取 1000 户，对每户中若干成员收视各电视节目的情况用仪器作详细记录。由此对各电视台、节目的收视情况作推断。

总体：全部上海市居民。个体：居民。

若用 X＝1 表示一个居民一周中看过某节目，X=0 表示未看过，则 X 就是一个变量，其总体分布可用 B(1,p)来表示。p 是指全部上海市民一周中看过该节目的人数比例。

若 Y 表示一个居民一周中看某节目的时间，则 Y 的总体分布可以用某种取值非负的随机变量的概率分布来描述。

样本(sample)：总体中的部分个体构成的集合。

样本容量(sample size)：样本中所含个体的数目。

抽样是统计研究中的最常用的一种手段。只要样本对总体有良好的代表性，那就可以对总体进行合理的推断。抽样方法对样本的代表性有重要影响。

通常采用**随机抽样**，其主要优点：

- 采用随机化机制抽样，不受主观意志的误导，样本代表性好。
- 能借助概率统计理论科学地推断总体信息，并对误差大小给出估计。
- 能根据精度要求、费用限制，事先确定合理的样本容量。

若采用随机方法抽取一个容量为 n 的样本，那么抽样之前哪些个体被抽中是不确定的，因此变量 X 的样本观测值也是随机的,记为 $X_1, X_2, \mathbf{L}, X_n$，其联合分布称为**样本的分布**。抽样之后，变量 X 的样本观测值确定下来了，记为 $x_1, x_2, \mathbf{L}, x_n$。样本的分布取决于总体分布与抽样方法，是统计推断的主要依据。

独立同分布(i.i.d.)的样本（也称**简单随机样本**）是数理统计中研究得最多的一种情况，满足如下两条性质：

- 代表性：样本每一个分量 X_i 的边际概率分布都与总体分布相同；
- 独立性：样本各分量 $X_1, X_2, \mathbf{L}, X_n$ 之间相互独立。

对于有限总体进行有放回、等概率的随机抽样得到的就是 i.i.d.样本；但是采取不放回抽样得到的就不是 i.i.d. 样本。对于无限总体，如果样本可视为在相同条件下重复观测获得，且各次观测之间互不影响，那么，i.i.d. 是对样本的一种合理假定。

i.i.d. 样本的分布结构较为简单。设总体分布函数为 $F(x;q)$，q 表示未知参数。那么 i.i.d. 样本 $(X_1, X_2, \mathbf{L}, X_n)$ 的联合分布函数为

$$F_{sample}(x_1, \mathbf{L}, x_n; q) = \prod_{i=1}^{n} F(x_i; q)。$$

若总体分布是连续型的或者离散型的，具有概率密度函数 $p(x;q)$，则 i.i.d. 样本的联合密度函数为

$$p_{sample}(x_1, \mathbf{L}, x_n; q) = \prod_{i=1}^{n} p(x_i; q)。$$

三、 描述性统计分析

在收集了数据（样本或普查数据）之后，首先要对数据进行适当的整理、汇总，可以帮助我们：

- 检查数据中是否有错漏；
- 概括数据的主要特征、揭示潜在结构（分布形状如何、变量间的相互关系如何、是否有离群点等）；
- 为进一步的分析提供启示。

常用方法：

- 图形：点线图、茎叶图、盒子图、直方图；条形图、饼图；散点图、时间序列图等。
- 表格：频率分布表、列联表等。
- 常用统计量：众数、中位数、均值；极差、方差、四分位间距；相关系数等。

注意：不同的方法适用于不同类型的数据。

四、 统计量与抽样分布

定义 7.1 设 $(X_1, X_2, \mathbf{L}, X_n)$ 是从某总体中抽取的一个容量为 n 的样本；$T(x_1, \mathbf{L}, x_n)$ 是一个实值或实向量值函数，不含任何未知参数，且定义域包含 $(X_1, X_2, \mathbf{L}, X_n)$ 的样本空间。则称 $T = T(X_1, X_2, \mathbf{L}\, X_n)$ 为**统计量**(statistic)，其概率分布称为**抽样分布**(sampling distribution)。

统计量就是概括、分析样本观测数据的方法。抽样之前，因样本的随机性，统计量也是随机的。通过抽样分布就能了解统计量所反映的总体信息，以及用之作推断时的效果好坏、误差大小。抽样分布的研究时数理统计学的一项重要内容。

例 7.2 设总体 $X \sim N(m, s^2)$，其中 m, s^2 是未知参数。从中抽取一个简单随机样本 $(X_1, \mathbf{L}, X_n)$，则 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，$T_1 = X_1 + X_2$，$T_2 = X_{(n)} - X_{(1)}$ 都是统计量。(T_1, T_2) 是一个二维的统计量。但 $\bar{X} - m$，$\frac{\bar{X} - m}{s}$ 不是统计量。由正态分布性质知 $\bar{X} \sim N(m, \frac{s^2}{n})$，这就是统计量 $\bar{X}$ 的抽样分布。

常用统计量及其抽样分布 设 $(X_1, \mathbf{L}, X_n)$ 是来自某总体的样本。

1. **样本均值** $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$。样本均值主要反映总体均值的信息。

定理7.1 样本均值的性质：

1. 若 $X_1, \mathbf{L}, X_n$ *i.i.d.* $N(m, s^2)$，则 $\bar{X} \sim N\left(m, \frac{s^2}{n}\right)$.

2. 若 $X_1, \mathbf{L}, X_n$ *i.i.d.* 总体 X，总体分布的均值为 m，方差为 s^2，则

(1) $E\bar{X} = m$，$Var(\bar{X}) = s^2/n$；

(2) 当样本容量 n 充分大时，$\bar{X}$ 近似 $N\left(m, \frac{s^2}{n}\right)$.

2. 样本方差

称 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 为**样本方差**，称 $S=\sqrt{S^2}$ 为**样本标准差**。称 $S_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 为**有偏方差**。样本方差主要反映总体方差的信息。

常用公式

$$Q=\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2=\sum_{i=1}^{n}\left(X_i-c\right)^2-n(\bar{X}-c)^2\ 。$$

定理7.2 样本方差的性质：

1. 若 $X_1,\mathbf{L},X_n$ *i.i.d.* 总体 X，总体分布的方差为 s^2，则 $E(S^2)=s^2$;

2. 若 $X_1,\mathbf{L},X_n$ *i.i.d.* $N(m,s^2)$，则

(1) $\bar{X}\sim N\left(m,\frac{s^2}{n}\right)$;

(2) $\frac{(n-1)S^2}{s^2}\sim c^2(n-1)$;

(3) $\bar{X}$ 与 S^2 相互独立;

(4) $T=\frac{\sqrt{n}\left(\bar{X}-m\right)}{S}\sim t(n-1).$

3. 样本矩 (sample moments)

样本各阶矩主要反映总体各阶矩的信息。

样本 v 阶原点矩 $A_v=\frac{1}{n}\sum_{i=1}^{n}X_i^v$，样本 v 阶中心矩 $B_v=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^v$ 。

定理7.3 (样本矩的性质) $X_1,\mathbf{L},X_n$ *i.i.d.* 总体 X。

(1) 若总体分布的 v 阶原点矩 m_v 存在，则 $E(A_v)=m_v$;

(2) 若 m_{2v} 存在，则 $Var(A_v)=\frac{1}{n}(m_{2v}-m_v^2).$

4. 样本偏态系数、峰态系数

样本偏态、峰态系数主要反映总体分布的偏态、峰态系数的信息。

称 $Sk=\dfrac{B_3}{B_2^{3/2}}=\dfrac{B_3}{\left(S_n^2\right)^{3/2}}$ 为样本偏态系数，

称 $Ku=\dfrac{B_4}{B_2^2}-3=\dfrac{B_4}{\left(S_n^2\right)^2}-3$ 为样本峰态系数。

5. 次序统计量 (order statistics)

设 $(X_1,\mathbf{L},X_n)$ 是从总体 X 中抽取一个容量为 n 的样本，将它们从小到大排序为 $X_{(1)}\le\mathbf{L}\le X_{(n)}$。称 $(X_{(1)},\mathbf{L},X_{(n)})$ 为样本的**次序统计量**；称 $X_{(i)}$ 为**第i个次序**统计量。

$X_{(1)}=\min\{X_1,X_2,\mathbf{L},X_n\}$ 称为最小次序统计量；

$X_{(n)}=\max\{X_1,X_2,\mathbf{L},X_n\}$ 称为最大次序统计量；

$R=X_{(n)}-X_{(1)}$ 称为样本极差(range)。

若总体分布已知，那么单个次序统计量、若干个次序统计量的联合分布都是可以求的。

关于连续型分布总体的次序统计量的抽样分布的主要结果：

设 $(X_1,X_2,\mathbf{L},X_n)$ 是取自总体 X 的一个简单随机样本，X 的分布函数是 $F(x)$，密度函数是 $p(x)$。则：

(1) 第 k 个次序统计量的 p.d.f. 为

$$p_k(x)=\binom{n}{k-1}\binom{n-k+1}{1}\left[F(x)\right]^{k-1}p(x)\left[1-F(x)\right]^{n-k}$$
$$=\frac{n!}{(k-1)!(n-k)!}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k}p(x)$$

(2) 第 j, 第 k 个次序统计量($j<k$)的 j.p.d.f. 为

$$p_{jk}(x_j,x_k)=$$
$$\begin{cases}\dfrac{n!}{(j-1)!(k-j-1)!(n-k)!}\left[F(x_j)\right]^{j-1}\left[F(x_k)-F(x_j)\right]^{k-j-1}\left[1-F(x_k)\right]^{n-k}p(x_j)p(x_k), & \text{if } x_j<x_k;\\ 0, & \text{else.}\end{cases}$$

样本极差的分布可由最大、最小次序统计量的联合分布推导得到。

例7.3 设 $X_1, \mathbf{L}, X_n$ $i.i.d.$ $Exp(l)$，则：

- $X_{(1)}$的pdf为 $p_1(x)=\begin{cases} n l e^{-n l x}, & x>0, \\ 0, & x\le 0. \end{cases}$

- $X_{(n)}$的pdf为 $p_n(x)=\begin{cases} n l (1-e^{-l x})^{n-1} e^{-l x}, & x>0, \\ 0, & x\le 0. \end{cases}$

- $(X_{(1)}, X_{(n)})$的jpdf为

$$p_{1,n}(x,y)=\begin{cases} n(n-1) l^2 e^{-l(x+y)} (e^{-l x}-e^{-l y})^{n-2}, & 0<x<y, \\ 0, & else. \end{cases}$$

- 极差 R 的pdf为

$$p_R(r)=\begin{cases} (n-1) l e^{-l r} (1-e^{-l r})^{n-2}, & r>0, \\ 0, & r\le 0. \end{cases}$$

注：

- 只有在少数几种总体分布下，少量几个统计量的精确分布能够用简洁的公式表示出来。
- 对于简单随机样本，当样本容量足够大时，由中心极限定理可得，样本矩等统计量渐近地服从正态分布。
- 有些情况下，可以用随机模拟得方法来寻找统计量的分布。

6. 正态总体下的三大抽样分布

- χ^2 分布
- t 分布
- F 分布

7. 经验分布函数

定义7.2 设从总体 X 中抽取一个容量为 n 的样本，样本观测值为 $x_1, x_2, \mathbf{L}, x_n$。将它们从小到大排序后记为 $x_{(1)}, x_{(2)}, \mathbf{L}, x_{(n)}$。对于 $x\in(-\infty,+\infty)$，令

$$F_n(x)=\begin{cases} 0, & x<x_{(1)}, \\ k/n, & x_{(k)}\le x<x_{(k+1)},\ k=1,2,\mathbf{L},n-1, \\ 1, & x\ge x_{(n)}, \end{cases}$$

称 $F_n(x)$ 为该样本的**经验分布函数**。

注：

- $F_n(x)$ = 观测值 $x_1, \mathbf{L}, x_n$ 中小于等于 x 的频率。
- $F_n(x)$ 是一个右连续、单调非降、取值于 $[0,1]$ 的阶梯函数，因而可视之为以等概率取 $x_1, \mathbf{L}, x_n$ 的离散型随机变量的分布函数。
- 抽样之前，对 $\forall x, F_n(x)$ 都是随机变量。

8. 格里汶科定理

定理7.4 设总体 X 的分布为 $F(x)$，$(X_1, X_2, \mathbf{L}, X_n)$ 是取自该总体的简单随机样本，对它的任意一个观测 数据 $(x_1, x_2, \mathbf{L}, x_n)$，记 $F_n(x)$ 为其经验分布函数，又记

$$D_n = \sup_{-\infty<x<+\infty} \left|F_n(x) - F(x)\right|.$$

则有

$$P(\lim_{n\to\infty} D_n = 0) = 1.$$

格里汶科定理说明，当样本容量足够大时，简单随机样本的任意一组观测数据的经验分布函数几乎都与总体分布函数充分接近。因此，由对样本数据的分布描述就可以得到总体分布的信息。

五、 充分统计量

定义 7.3 设总体分布族为 $\mathbf{F} = \{F_q : q \in \Theta\}$。$X_1, \mathbf{L}, X_n$ 是从总体中抽取的一个样本，$T = T(X_1, \mathbf{L}, X_n)$ 是一个统计量。若在给定 $T = t$ 的条件下，样本的条件分布 $(X_1, \mathbf{L}, X_n) \mid T = t$ 与 q 无关，对于 $\forall t$ 都成立。则称统计量 T 是分布族 $\mathbf{F}$ 的**充分统计量**。

注：

- 定义中的 q 可以是有限维参数，也可以是无限维参数。
- 在总体分布族假定下，充分统计量包含了样本中关于总体分布的所有信息。因而若用充分统计量替代原始样本作统计推断，没有任何损失。
- 充分性严重依赖于对总体分布族的假定。一个统计量对于 $\mathbf{F}_1$ 是充分统计量，但是对于 $\mathbf{F}_2$ 未必充分。
- 若统计量 T 是总体分布族 $\mathbf{F}$ 的充分统计量，$T = g(U)$，则 U 也是 $\mathbf{F}$ 的充分统计量。

例 7.4 设 $X_1, \mathbf{L}, X_n$ *i.i.d.* $B(1, p)$, $p \in (0,1)$. $n > 2$. 考察统计量

$$T_1 = \sum_{i=1}^{n} X_i, \quad T_2 = X_1 + X_2$$

的充分性。

样本的联合分布是

$$P(X_1 = x_1, \mathbf{L}, X_n = x_n) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i},$$

其中 $x_i = 0 \ or \ 1, \ i = 1, \mathbf{L}, n$。统计量 $T_1 \sim B(n, p)$。在给定 $T_1 = t$ 的条件下，样本的条件分布为

$$\begin{aligned} & P(X_1 = x_1, \mathbf{L}, X_n = x_n \mid T_1 = t) \\ &= \frac{P(X_1 = x_1, \mathbf{L}, X_n = x_n, T_1 = t)}{P(T_1 = t)} \\ &= \frac{P(X_1 = x_1, \mathbf{L}, X_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i)}{P(T_1 = t)} \\ &= \frac{p^t(1-p)^{n-t}}{\binom{n}{t} p^t(1-p)^{n-t}} = \binom{n}{t}^{-1} \end{aligned}$$

该条件分布与参数 p 无关，对任意 t 都成立。因此 T_1 是充分统计量。同理可验证，T_2 不是充分统计量。

定理 7.5 （因子分解定理）设样本的联合密度为 $p_q(\underset{\sim}{x}), \ q \in \Theta$。统计量 $T(\underset{\sim}{X})$ 是充分统计量，当且仅当存在非负函数 $g_q(T(\underset{\sim}{x}))$ 及与参数 q 无关的 $h(\underset{\sim}{x})$，使得

$$p_q(\underset{\sim}{x}) = g_q(T(\underset{\sim}{x}))h(\underset{\sim}{x}), \quad \forall \underset{\sim}{x}, \ \forall q \in \Theta.$$

例 7.5(1) 设 $X_1, \mathbf{L}, X_n$ 是来自于指数型分布族的 i.i.d. 样本。总体密度为

$$f(x;q) = c(q)\exp\left(\sum_{j=1}^k w_j(q)T_j(x)\right)h(x), \quad q \in \Theta,$$

其中 q 是 d 维参数，$d \le k$。则

$$T(X_1, \mathbf{L}, X_n) = \left(\sum_{i=1}^n T_1(X_i), \mathbf{L}, \sum_{i=1}^n T_k(X_i)\right)$$

是充分统计量。这个结果包含了二项分布、Poisson 分布、正态分布、指数分布、Gamma 分布等诸多分布族的情况。

(2) $X_1, \mathbf{L}, X_n \ \ i.i.d. U(0,q), \ q > 0.$ 则样本的联合密度为

$$p(x_1, \mathbf{L}, x_n; q) = q^{-n} I_{(0<X_{(n)}<q)}(x_1, \mathbf{L}, x_n)。$$

所以，$X_{(n)}$ 是充分统计量。

(3) 设总体分布的密度函数为 $p(x;q),\ q\in\Theta$，$X_1,\mathbf{L},X_n$ 是 i.i.d.样本，则次序统计量 $(X_{(1)},\mathbf{L},X_{(n)})$ 是充分统计量。

定义 7.4 设总体分布族为 $\mathrm{F}=\{F_q:q\in\Theta\}$，$T=T(X_1,\mathbf{L},X_n)$ 是充分统计量，且是任意一个充分统计量 $S(X_1,\mathbf{L},X_n)$ 的函数，则称 T 为**最小充分统计量**。

定义 7.5 若一个统计量的抽样分布不依赖于任何未知参数，则称该统计量为**附属**的(ancillary)。

定义 7.6 设总体的未知参数$q\in\Theta$，T 是一个统计量。若

$$E_q g(T)=0,\ \forall q\in\Theta\ \Rightarrow\ P_q(g(T)=0)=1,\ \forall q\in\Theta$$

则称 T 是**完备统计量**。

定理 7.6 （Basu 定理）若 T 是充分完备统计量，则 T 与任何附属统计量都独立。

第八章　参数估计

一、　参数估计简介

参数估计是一类重要的统计推断形式。目的是根据样本观测数据对反映总体分布特征或决定总体分布的一些未知参数作出估计。

待估参数一般是实数或实数向量，包括以下一些类型：

1) 决定总体分布的未知参数，这里记为 q。例如，假定总体服从 $N\left(m, s^2\right)$，其中 m，s^2 是未知参数，待估计。

2) 1) 中参数的函数 $g(q)$。例如，

- 假定总体 X～ $N\left(m, s^2\right)$，　$q=(m, s^2)$，对给定的 c，欲估计概率 P(X<c)，它等于 $\Phi(\frac{c-m}{s}) \triangleq g(q)$；
- 总体分布的各种未知的数字特征。例如：未知的总体均值、总体方差、p 分位数、偏度系数、变异系数等，二维总体的相关系数等等。

参数估计的两种形式：点估计、区间估计。

点估计 (point estimation)：

对未知参数 $g(q)$ (实值或实向量) 选用一个维数与之相同的统计量 $\hat{h}(X_1, \mathbf{L}, X_n)$ 来作推断。该统计量称为 $g(q)$ 的估计量 (estimator)。获得样本观测值后，代入估计量，就得到 $g(q)$ 的估计值 (estimate) $\hat{h}(x_1, \mathbf{L}, x_n)$。

区间估计 (interval estimation):

选用两个满足如下大小关系的统计量 $\hat{h}_L(X_1, \mathbf{L}, X_n) \le \hat{h}_U(X_1, \mathbf{L}, X_n)$ 构造一个随机区间来推断一维实值未知参数 $g(q)$。获得样本观测值后，代入此随机区间，就获得了概括 $g(q)$ 信息的一个具体的区间 $[\hat{h}_L(x_1, \mathbf{L}, x_n), \hat{h}_U(x_1, \mathbf{L}, x_n)]$。

问题：如何构造估计量？
如何评价估计量的表现？

二、 构造点估计量的两种常用方法

1．矩法估计 (Moment Estimation)

矩法估计源于英国统计学家 K. Pearson 提出的替换原则：

- 用样本矩去替换相应的总体矩（这里矩可以是原点矩也可以是中心矩）；
- 用样本矩的函数去替换相应的总体矩的函数。

设 $X_1, \mathbf{L}, X_n$ 是来自于总体 X 的简单随机样本，则按矩法估计的替换原则，有：

- 总体 m 阶原点矩 m_m 的估计量为：$\hat{m}_m = \frac{1}{n}\sum_{i=1}^{n} X_i^m,\ m = 1, 2, \mathbf{L}$；
- 总体 m 阶中心矩 n_m 的估计量为：$\hat{n}_m = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^m$；
- 总体均值 $m = EX$ 的估计量为：$\hat{m} = \bar{X}$；
- 总体方差 $s^2 = Var(X)$ 的估计量为：$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 = S_n^2$；
- 二维总体的相关系数 $r = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}}$ 的矩估计量是样本相关系数：

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}};$$

等等

求总体分布未知参数 $q_1, \mathbf{L}, q_k$ 的矩估计量的一般方法：

1) 求出总体分布的前 k 阶矩 $m_1, \mathbf{L}, m_k$ 关于 $q_1, \mathbf{L}, q_k$ 的函数

$$m_j = m_j(q_1, \mathbf{L}, q_k), \quad j = 1, 2, \mathbf{L}, k;$$

2) 求此函数的反函数

$$q_j = q_j(m_1, \mathbf{L}, m_k), \quad j = 1, 2, \mathbf{L}, k,$$

将其中的 m_j 用相应的样本矩 $\hat{m}_j$ 替代，即得 $q_1, \mathbf{L}, q_k$ 的矩估计量

$$\hat{q}_j = q_j(\hat{m}_1, \mathbf{L}, \hat{m}_k), \quad j = 1, 2, \mathbf{L}, k。$$

若待估参数为 $x = g(q_1, \mathbf{L}, q_k)$，则其中的 $q_1, \mathbf{L}, q_k$ 用其矩估计量 $\hat{q}_1, \mathbf{L}, \hat{q}_k$ 替代，即可得 x 的矩估计量 $\hat{x} = g(\hat{q}_1, \mathbf{L}, \hat{q}_k)$。

例8.1 设 $X_1, \mathbf{L}, X_k$是来自总体 X的简单随机样本。

1) 若 $X \sim N(m, s^2)$，其中m, s^2是未知参数。因为m、s^2分别是总体均值、方差，所以它们的矩估计量分别为 $\hat{m} = \bar{X}$，$\hat{s}^2 = S_n^2$。

2) 若 $X \sim Exp(l)$，其中 l是未知参数。因为 $EX = \frac{1}{l}$, 求此反函数得 $l = \frac{1}{EX}$,

所以 l 的矩估计量为 $\hat{l} = \frac{1}{\bar{X}}$; $h = P(X \geq c) = e^{-lc}$ 的矩估计量为$\hat{h} = e^{-c/\bar{X}}$.

3) 若 $X \sim U(a,b)$，其中a, b是未知参数。因为

$$EX = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}$$

求反函数可得

$$a = EX - \sqrt{3Var(X)}, \quad b = EX + \sqrt{3Var(X)},$$

所以a,b的矩估计量为

$$\hat{a} = \bar{X} - \sqrt{3S_n^2}, \quad \hat{b} = \bar{X} + \sqrt{3S_n^2}.$$

4) 若 $X \sim P(l)$。因为 $l = EX = Var(X)$，所以 l 的矩估计量可以为

$$\hat{l} = \bar{X} \quad \text{或者} \quad \hat{l} = S_n^2,$$

但通常取前者。

2. 极大似然估计 (Maximum Likelihood Estimation)

极大似然估计最早是由 Gauss 在 1821 年提出的，但一般将之归功于 R. A. Fisher，因为他在 1922 年重新提出这种想法，并证明了它的一些性质，使之得到广泛应用。

基本思想：

将样本观测视为“结果”，将总体分布（或决定总体分布的参数q ）视为产生“结果”的“原因”。在获得样本观测值后，看哪种“原因”产生该“结果”的可能性最大，然后用该“原因”对应的参数值作为q 的估计值。

似然函数 (likelihood function)

用于衡量取得某样本观测值的可能性大小。

1) 若总体 X服从离散型分布，分布列为

$$P(X = a_j) = p(a_j; q), \quad j = 1, 2, \mathbf{L},$$

其中q 是未知参数，设它的可能取值范围为Θ。因为样本 $X_1, \mathbf{L}, X_n$ i.i.d. X，所以获得样本观测值 $x_1, \mathbf{L}, x_n$ 的概率为：

$$P(X_1 = x_1, \mathbf{L}, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} p(x_i; q).$$

这个概率可看作是未知参数q 的函数，称之为样本观测值 $x_1, \mathbf{L}, x_n$ 的似然函数，记作

$$L(q; x_1, \mathbf{L}, x_n) = \prod_{i=1}^{n} p(x_i; q), \quad q \in \Theta.$$

2) 若总体分布是连续型的，p.d.f.为 $p(x;q)$，其中未知参数 $q \in \Theta$。则获得样本观测值 $x_1, \mathbf{L}, x_n$ 的可能性大小可用样本的联合密度函数在 $x_1, \mathbf{L}, x_n$ 处的值来度量：

$$f(x_1, \mathbf{L}, x_n; q) = \prod_{i=1}^{n} p(x_i; q).$$

将它看作未知参数 q 的函数，也称之为样本观测值 $x_1, \mathbf{L}, x_n$ 的似然函数，记作

$$L(q; x_1, \mathbf{L}, x_n) = \prod_{i=1}^{n} p(x_i; q), \quad q \in \Theta.$$

定义 8.1 对于每个样本观测值 $x_1, \mathbf{L}, x_n$，令 $\hat{q}(x_1, \mathbf{L}, x_n) = \arg\max_{q \in \Theta} L(q; x_1, \mathbf{L}, x_n)$，则称 $\hat{q}(X_1, \mathbf{L}, X_n)$ 为未知参数 q 的极大似然估计量(MLE)。

定理 8.1 (MLE 的不变原则) 若 $\hat{q}$ 是 q 的 MLE，则 q 的任一函数 $h = g(q)$ 的 MLE 为 $\hat{h} = g(\hat{q})$。

求极大似然估计的基本步骤：

1) 写出似然函数 $L(q; x_1, \mathbf{L}, x_n), q \in \Theta$;
2) 求似然函数的最大值点。

对数似然函数与似然函数的极大值点相同，因此，常用对数似然函数求MLE.
对于 i.i.d. 样本，

$$l(q) = \ln L(q; x_1, \mathbf{L}, x_n) = \sum_{i=1}^{n} \ln p(x_i; q), \quad q \in \Theta.$$

求导求极值法

若 $L(q)$ (或$l(q)$) 是 $q = (q_1, \mathbf{L}, q_k)$ 的凹函数，且二阶可微，则 $L(q)$ 的最大值点存在，可由求解下述似然方程（组）获得：

$$\frac{\partial l(q)}{\partial q_j} = 0, \quad j = 1, \mathbf{L}, k$$

或求解方程（组）

$$\frac{\partial L(q)}{\partial q_j} = 0, \quad j = 1, \mathbf{L}, k.$$

例8.2 (1) $X_1,\mathbf{L},X_n$ *i.i.d.* $B(1,p)$, $p\in[0,1]$, 求 p 的MLE。

解：样本观测值 $x_1,\mathbf{L},x_n$ 的似然函数为

$$L(p;x_1,\mathbf{L},x_n)=\prod_{i=1}^{n}\left\{p^{x_i}(1-p)^{(1-x_i)}\right\}=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i},\quad p\in[0,1],$$

对数似然函数为

$$l(p)=T\ln p+(n-T)\ln(1-p),\ \text{其中 } T=\sum_{i=1}^{n}x_i.$$

$l(p)$ 在 $(0,1)$ 上二阶可微、严凹，故存在唯一极大值点。似然方程为

$$\frac{dl(p)}{dp}=\frac{T}{p}-\frac{n-T}{1-p}=0,$$

解得 p 的MLE为

$$\hat{p}=\frac{T}{n}=\bar{X}.$$

例8.2 (2) $X_1,\mathbf{L},X_n$ *i.i.d.* $N(m,s^2)$, $(m,s^2)\in(-\infty,\infty)\times(0,\infty)$, 求 m,s^2 的MLE。

解：样本观测值 $x_1,\mathbf{L},x_n$ 的似然函数为

$$L(m,s^2;x_1,\mathbf{L},x_n)=\prod_{i=1}^{n}\left\{\frac{1}{\sqrt{2p}\,s}e^{-\frac{(x_i-m)^2}{2s^2}}\right\}=c\left(s^2\right)^{-\frac{n}{2}}e^{-\frac{1}{2s^2}\sum_{i=1}^{n}(x_i-m)^2},$$

对数似然函数为

$$\begin{aligned}l(m,s^2)&=-\frac{n}{2}\ln s^2-\frac{1}{2s^2}\sum_{i=1}^{n}\left(x_i-m\right)^2+c'\\&=-\frac{n}{2}\ln s^2-\frac{1}{2s^2}\left[\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+n\left(\bar{x}-m\right)^2\right]+c'.\end{aligned}$$

对于 $\forall s^2>0$, 当 $m=\bar{x}$ 时，$l(m,s^2)$ 达到最大。再求 $l(\bar{x},s^2)$ 的最大值点。

可证 $l(\bar{x},s^2)$ 关于 s^2 凹、二次可微，故其最大值点为下列方程的解：

$$\frac{dl(\bar{x},s^2)}{ds^2}=-\frac{n}{2s^2}+\frac{1}{2(s^2)^2}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=0,$$

即 $s^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$。所以 m,s^2 的MLE为 $\hat{m}=\bar{X},\quad \hat{s^2}=S_n^2.$

例8.2 (3) $X_1,\mathbf{L},X_n$ *i.i.d.* $P(l)$, $l>0$. 求 l 的MLE。

解：样本观测值 $x_1,\mathbf{L},x_n$ 的似然函数为

$$L(l;x_1,\mathbf{L},x_n)=\prod_{i=1}^{n}\left\{\frac{l^{x_i}}{x_i!}e^{-l}\right\}=e^{-nl}\,l^{\sum_{i=1}^{n}x_i}\cdot c,\quad l>0.$$

对数似然函数为

$$l(l)=-nl+T\ln l+c',\ \text{其中 } T=\sum_{i=1}^{n}x_i.$$

$l(l)$ 在 $(0,\infty)$ 上二阶可微、严凹，故存在唯一极大值点。似然方程为

$$\frac{dl(l)}{dl}=-n+\frac{T}{l}=0,$$

其解为 $l=T/n$。故 l 的MLE为 $\hat{l}=\bar{X}$。

不能求导求极值的问题

例8.3 $X_1, \mathbf{L}, X_n$ i.i.d. $U(0,q)$, $q>0$. 求q的MLE。

解：样本观测值 $x_1, \mathbf{L}, x_n$ 的似然函数为

$$L(q; x_1, \mathbf{L}, x_n) = \prod_{i=1}^{n} p(x_i; q) = \begin{cases} \dfrac{1}{q^n}, & x_{(n)} \le q, \\ 0, & \text{else.} \end{cases}$$

其最大值点为$q = x_{(n)}$，所以q的MLE为 $\hat{q} = X_{(n)}$.

根据不变原则求 MLE

例8.4 (1) $X_1, \mathbf{L}, X_n$ *i.i.d.* $X \sim Exp(\dfrac{1}{l})$, $l>0$. 求 $m = EX$ 的MLE。

解：样本观测值 $x_1, \mathbf{L}, x_n$ 的似然函数为

$$L(l; x_1, \mathbf{L}, x_n) = \prod_{i=1}^{n} \left\{ l e^{-l x_i} \right\} = l^n e^{-l \sum_{i=1}^{n} x_i}, \quad l>0.$$

对数似然函数为 $\qquad l(l) = n \ln l - l \sum_{i=1}^{n} x_i,$

$l(l)$在$(0,\infty)$上二阶可微、严凹，故存在唯一极大值点。由似然方程

$$\frac{dl(l)}{dl} = \frac{n}{l} - \sum_{i=1}^{n} x_i = 0,$$

解得 l 的MLE为 $\hat{l} = \dfrac{n}{\sum_{i=1}^{n} X_i}$。因为 $m = \dfrac{1}{l}$，所以 m 的MLE为$\hat{m} = \bar{X}$.

例8.4 (2) $X_1, \mathbf{L}, X_n$ *i.i.d.* $X \sim N(m, s^2)$, m, s^2是未知参数。

求 $l = P(X>3)$的MLE。

解：因为 $l = P(X>3) = 1 - \Phi\left(\dfrac{3-m}{s} \right)$，而 m, s^2的MLE为 $\bar{X}, S_n^2$。由不变原则知，

l 的MLE为 $\quad \hat{l} = 1 - \Phi\left(\dfrac{3-\bar{X}}{\sqrt{S_n^2}} \right).$

三、 评价估计量的优良性准则

对于同一个未知参数，往往可以构造多个不同的估计量去估计，那么哪一种估计更好呢？每个估计量都有误差，误差大小随样本观测值变而变。故不能以具体某一次的使用效果来评价估计量的优劣，而应当根据长期使用的效果去评价。

常用准则：

均方误差准则

无偏性(unbiasedness)

有效性(efficiency)

相合性(consisitency) 等大样本性质

1. 均方误差准则

用平均误差的大小去评价估计量的优劣是非常自然的。均方误差(Mean Squared Error, MSE) 是衡量“平均误差”的一种方法，数学处理较方便。

定义 8.2 设 $\hat{h}$ 是一维参数 $g(q)$ 的估计量，称 $\mathrm{E}_q(\hat{h}-g(q))^2$ 为 $\hat{h}$ 的均方误差，记为 $MSE_q(\hat{h})$。

显然，

$$MSE_q(\hat{h}) = E_q\left[(\hat{h}-E_q\hat{h})+(E_q\hat{h}-g(q))\right]^2 = Var_q(\hat{h})+bias_q^2(\hat{h}).$$

定义 8.3 设 $q\in\Theta$，$\hat{h}_1(X_1,\mathbf{L},X_n),\hat{h}_2(X_1,\mathbf{L},X_n)$ 是一维参数 $g(q)$ 的两个估计量，若有

$$\mathrm{MSE}_q\left[\hat{h}_1(X_1,\mathbf{L}\,X_n)\right]\le\mathrm{MSE}_q\left[\hat{h}_2(X_1,\mathbf{L}\,X_n)\right],\quad \forall q\in\Theta,$$

且 $\exists q_0\in\Theta$ 使不等式严格成立，则称在均方误差意义下 $\hat{h}_1$ 优于 $\hat{h}_2$。

例8.5 $X_1,\mathbf{L},X_n$ i.i.d. $N(m,s^2)$, m,s^2 未知，欲估计 s^2。已知 S_n^2 是 s^2 的有偏估计，S^2 是 s^2的无偏估计。那么哪个在均方意义下更优呢？

由于 $Y=\dfrac{nS_n^2}{s^2}=\dfrac{(n-1)S^2}{s^2}\sim c_{n-1}^2$，可计算得：

$$MSE(S_n^2)=E(S_n^2-s^2)^2=\frac{2n-1}{n^2}s^4,\quad MSE(S^2)=E(S^2-s^2)^2=\frac{2}{n-1}s^4,$$

所以，在均方误差意义下，S_n^2 优于 S^2，$\forall n>1.$ S_n^2 的偏倚

$$bias(S_n^2)=E(S_n^2)-s^2=-\frac{1}{n}s^2\to 0\ (\text{当}n\to\infty)。$$

2. 无偏性

定义 8.4 设$q \in \Theta$，若一维参数$g(q)$的估计量$\hat{h}(X_1, \mathbf{L}, X_n)$满足

$$E\left[\hat{h}(X_1, \mathbf{L}\ X_n)\right] = g(q), \quad \forall q \in \Theta,$$

则称 $\hat{h}$ 是 $g(q)$ 的无偏估计 (unbiased estimator, U.E.)。

注：

1) 有的参数不存在 U.E.;
2) 若有 U.E. 存在，一般 U.E. 不唯一；
3) 若 $\hat{h}$ 是h 的 U.E.，通常$h(\hat{h})$ 不是 $h(h)$ 的 U.E.，例如，样本标准差 S 不是总体标准差 s 的 U.E.。

定理 8.2 设 $X_1, \mathbf{L}, X_n$ 是来自总体 X 的 *i.i.d.* 样本。

1) 若总体的 k 阶原点矩 $m_k = EX^k$ 存在，则样本 k 阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 m_k 的U.E.;

2) 若总体方差 s^2 存在，则 S^2 是 s^2 的U.E.，但 S_n^2 不是 s^2 的U.E.。

一个未知参数的无偏估计量也可以构造出很多，根据均方误差准则，我们可以用方差大小去比较不同无偏估计的优劣。

定义 8.5 设$q \in \Theta$，$\hat{h}_1(X_1, \mathbf{L}, X_n)$, $\hat{h}_2(X_1, \mathbf{L}, X_n)$ 都是一维参数$g(q)$的无偏估计量。若有

$$Var_q\left[\hat{h}_1(X_1, \mathbf{L}\ X_n)\right] \le Var_q\left[\hat{h}_2(X_1, \mathbf{L}\ X_n)\right], \quad \forall q \in \Theta,$$

且 $\exists q_0 \in \Theta$ 使不等式严格成立，则称则 $\hat{h}_1$ 比 $\hat{h}_2$ 有效。

无偏估计量的标准差的估计称为**标准误**(standard error, s.e.)，常用于根据样本观测值推断估计值的误差大小。

例如：$\bar{X}$ 是总体平均 m 的*U.E.*，若总体方差 s^2 存在，则 $\mathrm{std}(\bar{X}) = \frac{s}{\sqrt{n}}$。若用 S 估计 s，则 $\bar{X}$的标准误为 $S/\sqrt{n}$.

例8.6 (1) $X_1, \mathbf{L}, X_n$ *i.i.d.* X，$EX = m$，$Var(X) = s^2$，m, s^2未知。欲估计 m。

- $\bar{X}$ 与 $\frac{1}{2}(X_1 + X_2)$ 都是 m 的*U.E.*，但

$$Var(\bar{X}) = \frac{1}{n}s^2, \quad Var\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}s^2,$$

故当 $n > 2$ 时，$\bar{X}$ 比 $\frac{1}{2}(X_1 + X_2)$ 有效。

- 更一般地，若 $a_i, i = 1, \mathbf{L}, n$ 满足 $\sum_{i=1}^{n} a_i = 1$，则称 $T = \sum_{i=1}^{n} a_i X_i$ 是 m 的线性无偏估计(LUE)。$\bar{X}$ 是 m 的LUE中的一个。在所有LUE中哪个最有效呢？

例8.6 (2) $X_1, \mathbf{L}, X_n$ i.i.d. $U(0, q)$，$q > 0$ 未知。则 $2\bar{X}$ 是 q 的矩估计量，无偏；

$T = \frac{n+1}{n}X_{(n)}$ 是由 q 的MLE构造的 q 的无偏估计。可证：当 $n > 1$ 时，T 比 $2\bar{X}$ 有效。

定义 8.6 设 $q \in \Theta$，h^* 是 $g(q)$ 的 U.E.。若对于 $g(q)$ 的任意一个 U.E. $\hat{h}$，都有 $\mathrm{var}_q(h^*) \le \mathrm{var}_q(\hat{h}), \ \forall q \in \Theta$，则称 h^* 是 $g(q)$ 的 UMVUE。

注：

有的参数存在 UMVUE，有的没有。

有时候即使有 UMVUE 存在，也很难求得。

定理 8.3 (Rao-Blackwell) 设 $q \in \Theta$，T 是充分统计量，h 是 $g(q)$ 的 U.E.。记 $f(T) = E(\hat{h} \mid T)$。则 $f(T)$ 是 $g(q)$ 的 U.E.，且 $\mathrm{var}_q(f(T)) \le \mathrm{var}_q(\hat{h}), \ \forall q \in \Theta$。

定理 8.4 设 $q \in \Theta$，h 是 $g(q)$ 的 U.E.。h 是 $g(q)$ 的 UMVUE 的充要条件是：h 与所有方差有限的 0 的无偏估计不相关。

定理 8.5 (Lehmann-Scheffé) 设 $q \in \Theta$，T 是充分完备统计量，$f(T)$ 是一个统计量，且是 T 的函数，则 $f(T)$ 是其期望 $g(q) = E_q f(T)$ 的 UMVUE。

3．相合性 (consistency)

定义 8.6 设$q \in \Theta$，若一维参数$g(q)$的估计量$\hat{h}(X_1, \mathbf{L}, X_n)$满足：对$\forall e > 0$，有

$$\lim_{n\to\infty} P_q\left[|\hat{h} - h| \geq e\right] = 0, \quad \forall q \in \Theta,$$

则称 $\hat{h}(X_1, \mathbf{L}, X_n)$ 是$g(q)$的相合估计。

注：相合性的意思是，随着样本容量增大，估计量应当越来越接近参数的真实值。这是对估计量的一个最基本的要求。

判断相合性的基本工具：

1) 大数定律；

2) 看是否有 $\lim_{n\to\infty} MSE(\hat{h}_n) = 0;$ （因为$P\left(|\hat{h}_n - h| \geq e\right) \leq \frac{MSE(\hat{h}_n)}{e^2}$）

3) 若$\hat{h}_n^1, \mathbf{L}, \hat{h}_n^k$分别是$h_1, \mathbf{L}, h_k$的相合估计，$x = g\left(h_1, \mathbf{L}, h_k\right)$是连续函数，则$g(\hat{h}_n^1, \mathbf{L}, \hat{h}_n^k)$是$x$的相合估计。

例8.7 1) $X_1, \mathbf{L}, X_n$ *i.i.d.* X，若$m_k = EX^k$存在，则A_k是m_k的相合估计。

2) $X_1, \mathbf{L}, X_n$ *i.i.d.* $N(m, s^2)$. 因为$\lim_{n\to\infty} MSE(S_n^2) = \lim_{n\to\infty} MSE(S^2) = 0,$

所以S_n^2、S^2都是s^2的相合估计。$\bar{X}$是m的相合估计，但$\frac{1}{2}(X_1 + X_2)$不是。

极大似然估计的渐近正态性

定理 8.6 设 $\mathrm{F} = \{p(x;q) : q \in \Theta\}$ 是一个概率密度族，其中Θ是 ¡ 上的非退化区间。假如该分布族满足下列正则条件：

1) $\forall q \in \Theta,$ 偏导数 $\frac{\partial \log p}{\partial q}, \frac{\partial^2 \log p}{\partial q^2}, \frac{\partial^3 \log p}{\partial q^3}$ 存在；

2) $\forall q \in \Theta,$ 有 $\left|\frac{\partial p}{\partial q}\right| < g_1(x), \left|\frac{\partial^2 p}{\partial q^2}\right| < g_2(x), \left|\frac{\partial^3 \log p}{\partial q^3}\right| < H(x)$，其中 $g_1(x), g_2(x)$ 在实数轴上可积，且 $H(x)$ 满足 $\int_{-\infty}^{\infty} H(x)p(x;q)dx < M$， M 与 q 无关。

3) $\forall q \in \Theta,$ 有 $0 < E\left(\frac{\partial \log p}{\partial q}\right)^2 = \int_{-\infty}^{\infty}\left(\frac{\partial \log p}{\partial q}\right)^2 p(x;q)dx < \infty$。

则在分布参数 q 的未知真值q_0为Θ的一个内点的情况下，其似然方程必有一个解依概率收

敛于真值 q_0，且渐近地服从正态分布 $N\left(q_0,\left[nE\left(\frac{\partial \log p}{\partial q}\right)^2\Big|_{q=q_0}\right]^{-1}\right)$。

对于估计量的优劣可以从多个角度评价。有的估计量从某种角度看较好，但从别的角度看未必好。可以说，几乎没有面面俱到的、各方面都好的估计量。但有从各方面来看都不好的估计量，这样的估计量显然没有任何使用价值。

四、区间估计

基本想法：

选用两个满足如下大小关系的统计量 $\hat{h}_L(X_1,\mathbf{L},X_n)\le\hat{h}_U(X_1,\mathbf{L},X_n)$ 构造一个随机区间来推断一维实值未知参数 $g(q)$。获得样本观测值后，代入此随机区间，就获得了概括 $g(q)$ 信息的一个具体的区间 $[\hat{h}_L(x_1,\mathbf{L},x_n),\hat{h}_U(x_1,\mathbf{L},x_n)]$。

对 $\hat{h}_L,\hat{h}_U$ 的自然要求：

1) 区间$[\hat{h}_L,\hat{h}_U]$中包含 h 的可能性尽量大；

2)区间长度尽量小。

定义8.7 设总体参数$q\in\Theta$，$g(q)$是一维参数，$X_1,\mathbf{L},X_n$是样本，$\hat{h}_L(X_1,\mathbf{L},X_n)$，$\hat{h}_U(X_1,\mathbf{L},X_n)$是两个统计量。若对给定的$a\in(0,1)$，有

$$P_q\left[\hat{h}_L(X_1,\mathbf{L},X_n)\le g(q)\le\hat{h}_U(X_1,\mathbf{L},X_n)\right]\ge 1-a,\ \forall q\in\Theta,$$

则称$[\hat{h}_L,\hat{h}_U]$是$\hat{h}$的置信水平为$1-a$的**置信区间**(Confidence Interval, C.I.)。

例 8.8 正态总体均值的置信区间。

1. $X_1,\mathbf{L},X_n$ *i.i.d.* $X\sim N(m,s^2)$，s^2已知，求m的置信区间。

   Q $\dfrac{\bar{X}-m}{s/\sqrt{n}}\sim N(0,1)$　　$\therefore\ P\left\{\left|\dfrac{\bar{X}-m}{s/\sqrt{n}}\right|\le z_{1-a/2}\right\}=1-a,$

   其中 $z_{1-a/2}$ 为 $N(0,1)$的**1-a/2**下侧分位点。

   $\therefore\quad P\left\{\bar{X}-z_{1-a/2}\dfrac{s}{\sqrt{n}}\le m\le\bar{X}+z_{1-a/2}\dfrac{s}{\sqrt{n}}\right\}=1-a,\ \forall m$

   因此，

   $$\left[\bar{X}-z_{1-a/2}\frac{s}{\sqrt{n}},\ \bar{X}+z_{1-a/2}\frac{s}{\sqrt{n}}\right]$$

   是 m 的置信水平为$1-a$的C.I.

2. $X_1, \mathbf{L}, X_n$ *i.i.d.* $X \sim N(m, s^2)$, m, s^2均未知，求 m 的置信区间。

Q $\dfrac{\bar{X}-m}{S/\sqrt{n}} \sim t(n-1)$ $\quad\therefore\ P\left\{\left|\dfrac{\bar{X}-m}{S/\sqrt{n}}\right| \le t_{1-\frac{a}{2}}(n-1)\right\} = 1-a,$

其中 $t_{1-\frac{a}{2}}(n-1)$ 为 $t(n-1)$ 的 $1-a/2$ 下侧分位点。

$$\therefore\ P\left\{\bar{X} - t_{1-\frac{a}{2}}(n-1)\frac{S}{\sqrt{n}} \le m \le \bar{X} + t_{1-\frac{a}{2}}(n-1)\frac{S}{\sqrt{n}}\right\} = 1-a,\ \forall m, s$$

所以

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

是 m 的置信水平为$1-a$ 的C.I.。

构造 C.I.的一般方法——枢轴量法

设欲构造未知参数 $g(q)$ 的CI。

1. 构造一个样本与 $g(q)$ 的函数 $G(X_1, \mathbf{L}, X_n; g(q))$，对$\forall q \in \Theta$，其分布完全已知。称此函数为**枢轴量** (pivotal quantity)。

   注：G不是统计量，它往往是由 $g(q)$ 的估计量构造的，刻划的是估计量与$g(q)$的某种绝对或相对的差异。

2. 选择两个常数 c, d，使对给定的 a 有 $\quad P(c \le G \le d) \ge 1-a$ (或 $=1-a$)。

3. 若能将不等式 $c \le G \le d$ 变形为 $\hat{h}_L \le g(q) \le \hat{h}_U$, 其中$\hat{h}_L, \hat{h}_U$是统计量，则得

$$P(\hat{h}_L \le g(q) \le \hat{h}_U) \ge (\text{或} =)\ 1-a, \quad \forall q \in \Theta.$$

注：2.中 c, d 的取法很多，理论上应取使 $E(\hat{h}_U - \hat{h}_L)$ 一致最小的 c, d，但大多数场合难以做到。因此常取满足下式的 c, d：

$$P(G < c) = P(G > d) = a/2.$$

例 8.9 正态总体方差的置信区间。$X_1, \mathbf{L}, X_n$ *i.i.d.* $X \sim N(m, s^2)$，m, s^2均未知，求 s^2 的置信区间。

取枢轴量为 $\quad G = \dfrac{(n-1)S^2}{s^2}$，$\quad G \sim c^2(n-1).$

取 $c = c^2_{a/2}(n-1),\ d = c^2_{1-a/2}(n-1),$ 则 $\quad P\{c \le G \le d\} = 1-a, \quad \forall m, s^2.$

解不等式 $\quad c \le G \le d$ 得 s^2 的$1-a$ C.I.

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

例 8.10 两正态总体均值之差的置信区间。

$X_1,\mathbf{L},X_m$ *i.i.d.* $N(m_1,s_1^2)$, $Y_1,\mathbf{L},Y_n$ *i.i.d.* $N(m_2,s_2^2)$, 两样本相互独立。求 m_1-m_2 的**CI**。

显然，$\bar{X}-\bar{Y}$ 是 m_1-m_2 的合适的估计量。可证

$$\bar{X}-\bar{Y}\sim N\left(m_1-m_2,\frac{s_1^2}{m}+\frac{s_2^2}{n}\right).$$

⑴ 若 s_1^2,s_2^2 已知，则可取枢轴量为

$$G=\frac{(\bar{X}-\bar{Y})-(m_1-m_2)}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}}.$$

$G\sim N(0,1)$, 故得 m_1-m_2 的 $1-a$ C.I. $\quad \bar{X}-\bar{Y}\pm z_{1-a/2}\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}.$

（2）若 $s_1^2=s_2^2=s^2$ 未知，则

$$S^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}$$

是 s^2 的合理估计。枢轴量可取为

$$G=\frac{(\bar{X}-\bar{Y})-(m_1-m_2)}{S\sqrt{\frac{1}{m}+\frac{1}{n}}}.$$

$G\sim t(m+n-2)$, 故得 m_1-m_2 的 $1-a$ C.I. $\quad \bar{X}-\bar{Y}\pm t_{1-a/2}(m+n-2)S\sqrt{\frac{1}{m}+\frac{1}{n}}.$

(3) 当 s_1^2、s_2^2 都未知时，构造 m_1-m_2 的置信区间的问题是历史上著名的 **Behrens-Fisher** 问题**(1929).**

当 m、n都很大时，由

$$\frac{\bar{X}-\bar{Y}-(m_1-m_2)}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}}\ \overset{\text{近似}}{\sim}\ N(0,1),$$

可得到 m_1-m_2 的置信水平近似为 $1-a$ 的置信区间

$$\left[\bar{X}-\bar{Y}-z_{1-a/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}},\ \bar{X}-\bar{Y}+z_{1-a/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}\right].$$

例 8.11 两正态总体方差之比的置信区间。

$X_1,\mathbf{L},X_m$ *i.i.d.* $N(m_1,s_1^2)$, $Y_1,\mathbf{L},Y_n$ *i.i.d.* $N(m_2,s_2^2)$, 两样本相互独立。求 $q=s_1^2/s_2^2$ 的置信区间。

由于 S_1^2、S_2^2 分别是 s_1^2、s_2^2 的合适的估计量，且

$$(m-1)S_1^2/s_1^2 \sim c^2(m-1), \quad (n-1)S_2^2/s_2^2 \sim c^2(n-1),$$

两者相互独立，因此

$$F=\frac{S_1^2/s_1^2}{S_2^2/s_2^2} \sim F(m-1,n-1).$$

取F作枢轴量，可得$q=s_1^2/s_2^2$的置信水平为$1-a$的C.I.

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

构造区间估计的大样本方法

有些问题中，很难求得枢轴量在小样本情况下的精确分布，但容易求得大样本情况下的近似分布。此时，可用该近似分布构造区间估计。

基本工具：中心极限定理。

设 $X_1, X_2, \mathbf{L}\ i.i.d., E(X_1)=m,\ Var(X_1)=s^2$, 则当$n\to\infty$时 $\dfrac{\bar{X}-m}{\sqrt{s^2/n}}$ 依 $N(0,1)$.

即 $$P\left(\left|\frac{\bar{X}-m}{s/\sqrt{n}}\right| \le z_{1-a/2}\right) \approx 1-a.$$

于是可得 m 的置信水平近似为$1-a$ 的CI $\left[\bar{X}-z_{1-a/2}\dfrac{s}{\sqrt{n}},\ \bar{X}+z_{1-a/2}\dfrac{s}{\sqrt{n}}\right]$.

若 s 未知，则可用其相合估计 S 替代，得CI为 $\left[\bar{X}-z_{1-a/2}\dfrac{S}{\sqrt{n}},\ \bar{X}+z_{1-a/2}\dfrac{S}{\sqrt{n}}\right]$.

例 8.12

1. 设 $X_1, X_2, \mathbf{L}, X_n\ i.i.d.\ B(1,p)$, 求 p 的置信区间。

记 $Z=\dfrac{\bar{X}-p}{\sqrt{\frac{1}{n}p(1-p)}}$. 当$n\to\infty$时，$Z$ 依 $N(0,1)$.

所以取 Z 作为枢轴量。由 $P(-z_{1-a/2} \le Z \le z_{1-a/2}) \approx 1-a$ 得p的近似$1-a$ C.I.

$$\frac{1}{1+\frac{l}{n}}\left[\bar{X}+\frac{l}{2n}-z_{1-\frac{a}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{l}{4n^2}},\ \bar{X}+\frac{l}{2n}+z_{1-\frac{a}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{l}{4n^2}}\right],$$

其中 $l=z_{1-a/2}^2$.

当 n 较大时，

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{1}{n}\bar{X}(1-\bar{X})}} \xrightarrow{\text{近似}} N(0,1).$$

也可取 Z 作为枢轴量，得 p 的另一种近似 $1-\alpha$ C.I.

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

(在前一个置信区间中略去 $\frac{1}{n}$ 的项，亦可得此C.I.。)

2. 设 $X_1, X_2, \mathbf{L}, X_m$ $i.i.d.$ $B(1, p_1)$, $Y_1, Y_2, \mathbf{L}, Y_n$ $i.i.d.$ $B(1, p_2)$, 且两样本相互独立。求 $p_1 - p_2$ 的置信区间。

因为当 $m, n \to \infty$ 时

$$\frac{\bar{X} - \bar{Y} - (p_1 - p_2)}{\sqrt{\frac{1}{m}\bar{X}(1-\bar{X}) + \frac{1}{n}\bar{Y}(1-\bar{Y})}} \xrightarrow{\text{近似}} N(0,1).$$

由此可得 $p_1 - p_2$ 的近似 $1-\alpha$ $C.I.$

$$\left[\bar{X} - \bar{Y} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{m} + \frac{\bar{Y}(1-\bar{Y})}{n}},\ \bar{X} - \bar{Y} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{m} + \frac{\bar{Y}(1-\bar{Y})}{n}}\right].$$

3. 设 $X_1, X_2, \mathbf{L}, X_n$ $i.i.d.$ $\mathrm{Poi}(l)$, 求 l 的置信区间。

因为当 $n \to \infty$ 时

$$\frac{\bar{X} - l}{\sqrt{l/n}} \xrightarrow{\text{近似}} N(0,1).$$

由此可得 l 的近似 $1-\alpha$ C.I.

$$\left[\bar{X} + \frac{l}{2n} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n} + \frac{l}{4n^2}},\ \bar{X} + \frac{l}{2n} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n} + \frac{l}{4n^2}}\right],$$

其中 $l = z_{1-\alpha/2}^2$. 略去其中 $\frac{1}{n}$ 的项可得近似 $1-\alpha$ C.I.

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}},\ \bar{X} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}\right].$$

例 8.13 某厂生产的某种轮胎的寿命 X 被认为服从 N(m, s^2)。现随机抽取 12 个产品，测得其寿命分别为(单位：万公里)：

4.68　4.85　4.32　4.85　4.61　5.02　5.20　4.60　4.58　4.72　4.38　4.70。

求这种轮胎的平均寿命 m 的置信水平为 0.95 的置信区间。

解：方差未知时，正态总体 m 的$1-a$ C.I.为：

$$[\bar{X} - t_{1-\frac{a}{2}}(n-1)\frac{S}{\sqrt{n}}, \bar{X} + t_{1-\frac{a}{2}}(n-1)\frac{S}{\sqrt{n}}].$$

这里，$a = 0.05,\ n = 12,$ 查表知 $t_{0.975}(11) = 2.2010$。由样本观测数据计算得

$$\bar{X} = 4.7092, \quad S = \sqrt{0.0615}.$$

故 m 的 0.95 C.I. 为

$$4.7092 \pm 2.2010\sqrt{0.0615}/\sqrt{12} = [4.5516, 4.8668].$$

例8.14 从全市抽取400户家庭，调查去年的年收入。发现收入低于贫困线的家庭数为**12**户。请对全市贫困家庭比例 p 给出**95%C.I.**。

解：记 $X_i = \begin{cases} 1, & \text{第 } i \text{ 户为贫困家庭}, \\ 0, & \text{否则}, \end{cases} \quad i = 1, \cdots, n,\ n = 400.$

根据题目条件有 $X_1, \cdots, X_n$ *i.i.d.* $B(1, p)$。p 的近似95% C.I.为

$$\bar{X} \pm z_{1-a/2}\sqrt{\bar{X}(1-\bar{X})/n}.$$

根据样本数据知，$\hat{p} = \bar{X} = 0.03$。故 p 的近似95%**C.I.**为 (0.013283, 0.046717).

第九章　假设检验

一、 基本概念

假设检验 (Hypothesis Testing) 是又一种重要的统计推断形式。

与参数估计相比：

- 参数估计：对未知的总体参数给出一个具体的数值估计；
- 假设检验：对涉及总体未知信息的某个命题作出拒绝或接受的判断。

这类统计推断有丰富的实际背景，例如：

- 彩票的开奖机制是否公平？
- 医生常常得根据血样化验报告判断就诊者是否患某种疾病；
- 试制新药时，药厂得根据小鼠试验得结果判断：新药是否疗效更好？是否有毒性等副作用？
- 工厂需要根据在线抽样检查得结果判断生产线是否运行正常；
- 产品交验时，买卖双方需根据抽样检验的结果判断，一批产品的合格率、平均寿命等指标是否达到合同要求；

等等

例 9.1 某厂生产 10 欧姆的电阻。根据以往的情况，可以认为该厂生产的这种类型的电阻实际阻值 $X \sim N(m, 0.1^2)$。现在随机抽取 10 个电阻，测得它们的实际阻值为：

9.9　10.1　10.2　9.7　9.9　9.9　10　10.5　10.1　10.2

该厂生产的这种类型的电阻，其实际平均阻值 m 是否确实等于标定值 10 欧姆？

这里 m 是未知参数，需回答的问题是：m=10 还是 $m\neq10$?
命题“m=10” 就是涉及 m 的一个**假设**，需根据样本信息去判断。
命题“$m\neq10$” 是另一个假设，是“m=10” 的**对立假设**。
假设检验就是要在这两个对立的假设中作出一个选择。

假设：关于总体分布（或未知参数）的命题。

解决假设检验问题的基本步骤

1. 对实际问题建模，形成假设检验问题。

将实际问题转化为一个统计问题，把欲判断的命题转化为 “假设”。
假设常常成对出现：一个称为**原假设**(null hypothesis)，常用 H_0 表示；另一个是它的对立面，称为**备择假设**(alternative hypothesis)，常用 H_1 表示。将它们联立起来就表示一个假设检验问题。

例：回答例 9.1 的问题，可转化为解决如下假设检验问题：

$$H_0: m=10 \quad v.s. \quad H_1: m\neq 10$$

注：

- 在统计假设检验中，H_0 与 H_1 的地位有所不同。

  H_0 通常是比较可靠的、若无充分证据不能轻易拒绝的假设。H_1 表示 H_0 不真时最可能的情况。

- 假设分：简单假设、复合假设

2. 选择检验统计量，给出拒绝域的形式

解决假设检验问题，就是要给出根据样本观测值作出接受 H_0 或拒绝 H_0 判断的规则。

也即将样本的可能取值范围（样本空间）一分为二，一个区域为**拒绝域(Region of Rejection，**记作 W)，另一个区域称为**接受域**。当样本观测落在 W 中时，作出拒绝 H_0 的判断，反之，则作出接受 H_0 的判断。这样的规则就称为检验(test)。

通常检验是通过一个**检验统计量**来构造的。该统计量应当集中了样本中与假设有关的尽可能多的信息。

例9.1中，检验的是关于总体均值 m 的假设，而样本中关于 m 的信息主要集中在样本均值 $\bar{X}$ 中，所以常用 $\bar{X}$ 作为检验统计量。因为 m 大时 $\bar{X}$ 大的可能性大，m 小时 $\bar{X}$ 小的可能性大，所以检验$H_0: m=10 \ \ v.s. \ \ H_1: m\neq 10$的直观判断规则是：

$$\text{当} |\bar{X}-10| \geq c \text{ 时拒绝 } H_0,$$

其中 c 是一个合适的常数。

该规则对应的拒绝域为 $W=\{(x_1,\mathbf{L},x_n): |\bar{X}-10| \geq c\}$.

注：统计学中常强调“拒绝域”而较少提“接受域”，原因在于否定一个命题容易，只需要举一个反例即可；而证实一个命题难。

“接受原假设”的实际含义是没有充分证据反对原假设，并不表示已证实了原假设。

3. 确定显著性水平

- 假设检验也是用样本信息去推断总体信息的一种形式，因为样本信息是不完全的，所以检验难免会犯错误。
- 假设检验中会出现的**两类错误**：

样本情况	判断	总体情况	
		H_0 真	H_1 真
$(x_1,\cdots,x_n)\in W$	拒绝 H_0	犯第 I 类错误（拒真）	√
$(x_1,\cdots,x_n)\notin W$	接受 H_0	√	犯第 II 类错误（受伪）

例 9.1 中的两类错误

考虑按前面的判断规则“当$|\bar{X}-10|\geq c$时拒绝H_0”或相应的拒绝域W来检验假设

$$H_0: m=10 \quad v.s. \quad H_1: m\neq 10$$

因为$\bar{X}\sim N(m,\frac{1}{10}0.1^2)$，所以不管$c$取何正值：

H_0 真时，样本总有可能落入拒绝域；

H_1 真时，样本同样也总有可能落入接受域。

犯两类错误的概率

一般地，设W为假设检验问题 $H_0: q\in\Theta_0 \quad v.s. \quad H_1: q\in\Theta_1$ 的拒绝域，则 W 相应的检验规则犯第I类错误的概率为：

$$P_q\{(X_1,\mathbf{L},X_n)\in W\},\quad q\in\Theta_0.$$

犯第II类错误的概率为：

$$1-P_q\{(X_1,\mathbf{L},X_n)\in W\},\quad q\in\Theta_1.$$

称

$$g_W(q)=P_q\{(X_1,\mathbf{L},X_n)\in W\},\quad q\in\Theta_0\cup\Theta_1$$

为检验规则W的**势函数**(power function).

例 9.1 中的势函数、犯两类错误的概率

势函数为：

$$\begin{aligned} g_c(m)&=P_m\left(|\bar{X}-10|\geq c\right)\\ &=1-P_m\left(\frac{10-c-m}{0.1/\sqrt{10}}<\frac{\bar{X}-m}{0.1/\sqrt{10}}<\frac{10+c-m}{0.1/\sqrt{10}}\right)\\ &=1-\left[\Phi\left(\frac{10+c-m}{0.1/\sqrt{10}}\right)-\Phi\left(\frac{10-c-m}{0.1/\sqrt{10}}\right)\right],\qquad m\in(-\infty,\infty). \end{aligned}$$

犯第I类错误的概率：$g_c(10)=2\left[1-\Phi(\frac{\sqrt{10}\cdot c}{0.1})\right]$；

犯第II类错误的概率：$1-g_c(m),\ m\neq 10.$

理想地，我们希望建立的检验规则能使犯两类错误的概率同时足够小。但这在样本容量固定的情况下做不到！所以只能折衷。

如例9.1中，$g_c(10)=2\left[1-\Phi(\frac{\sqrt{10}\cdot c}{0.1})\right]\downarrow$，若$c\uparrow$；但对任意给定的$m\neq 10$，$1-g_c(m)\uparrow$，若$c\uparrow$.

Neyman-Pearson 原则：在控制犯第 I 类错误概率的前提下，最小化犯第 II 类错误的概率。

显著性水平(level of significance)

对于假设检验问题 $H_0:q\in\Theta_0$ *v.s.* $H_1:q\in\Theta_1$，若一个检验规则(拒绝域为W)满足

$$\max_{q\in\Theta_0} g_W(q)\le a,$$

则称该检验规则的显著性水平为 a 。

注： a 的直观意义："小概率事件"或"不太可能事件"的标准。

a 的选择应视具体问题而定。

a 常取为 0.05，0.01，或 0.1。

按照 N-P 原则，在确保显著性水平的前提下，应选择使用犯第 II 类错误概率尽量小的检验规则。

4. 给出具体的拒绝域或检验规则

根据前面确定的检验统计量与拒绝域的形式、显著性水平 a，就可以给出具体的拒绝域或检验规则了。

例9.1中，若规定显著性水平为 a，则应取 c，使

$$\max_{m=10} g_c(m)=2\left[1-\Phi(10\sqrt{10}\cdot c)\right]\le a,$$

故 $c\ge z_{1-a/2}0.1/\sqrt{10}$，其中 $z_{1-a/2}$ 为标准正态分布的$1-a/2$分位点。同时，为最小化犯第二类错误的概率，c应尽可能地小，因此

$$c=z_{1-a/2}0.1/\sqrt{10}.$$

相应的检验规则也可表达为：

当 $\dfrac{|\bar{X}-10|}{0.1/\sqrt{10}}\ge z_{1-a/2}$ 时，拒绝 H_0.

5. 解决实际问题

例9.1中，取 $a=0.05$，则 $z_{0.975}=1.96$。由样本数据计算得：

$$\frac{|\bar{X}-10|}{0.1/\sqrt{10}}=\frac{|10.05-10|}{0.1/\sqrt{10}}=1.58114<z_{0.975}$$

所以，在0.05的显著性水平下不能拒绝原假设。即可认为该类电阻实际平均阻值等于标定值10欧姆。

对于一维的总体参数q,通常研究三种形式的假设检验问题：

(1) $H_0: q \le q_0 \quad v.s. \quad H_1: q > q_0$

(2) $H_0: q \ge q_0 \quad v.s. \quad H_1: q < q_0$

(3) $H_0: q = q_0 \quad v.s. \quad H_1: q \ne q_0$

(1)、(2)称为单边检验(one-side),(3)称为双边检验(two-side)。解决这些检验问题的检验统计量往往是相同的，差别在于拒绝域的形式不同。

二、构造检验的方法

1. 直观分析法

根据点估计量、充分性原则等依据选择合适的检验统计量，根据直观分析确定拒绝域的形式,然后通过计算势函数、分析临界值与犯两类错误概率的关系给出具体的检验规则。如例 9.1。

例 9.2 设样本 $X_1, \mathbf{L}, X_n \ i.i.d. N(m, s^2)$， s^2 已知，欲检验

$$H_0: m \ge m_0 \quad v.s. \quad H_1: m < m_0,$$

其中 m_0 是给定常数。

因 m 的信息主要集中于样本均值 $\bar{X}$ 中，故选 $\bar{X}$ 作为检验统计量。直观地，m 越大 $\bar{X}$ 也越大，m 越小 $\bar{X}$ 也越小，故拒绝域形式应取为 $\{\underset{\sim}{x}: \bar{X} \le c\}$，$c$ 是待定的临界值。该拒绝域的势函数为：

$$g_c(m) = P_m(\bar{X} \le c) = \Phi\left(\frac{c-m}{s/\sqrt{n}}\right),$$

它关于 m 单调降，关于 c 单调增。犯第一类错误的概率上限为

$$\max_{m \ge m_0} g_c(m) = g_c(m_0)。$$

若取显著性水平为 a ，则应取 c，使得 $g_c(m_0) \le a$ ，即 $c \le z_a s/\sqrt{n} + m_0$。而 c 越小，犯第二类错误的概率越大，故取 $c = z_a s/\sqrt{n} + m_0$。这样就得到了水平为 a 的拒绝域

$$\{\underset{\sim}{x}: \bar{X} \le z_a \frac{s}{\sqrt{n}} + m_0\} = \{\underset{\sim}{x}: Z \triangleq \frac{\bar{X} - m_0}{s/\sqrt{n}} \le z_a\}。$$

例 9.3 设样本 $X_1, \mathbf{L}, X_n$ $i.i.d.B(1, p)$，欲检验

$$H_0 : p \le p_0 \quad v.s. \quad H_1 : p > p_0 ,$$

其中 p_0 是给定常数。

因 $T = \sum_{i=1}^{n} X_i$ 为充分统计量，p 的信息主要集中在 T 中。直观地，取拒绝域的形式为 $\{\underset{\sim}{x} : T \ge c\}$，$c$ 是待定的临界值。因 $T \sim B(n, p)$，该拒绝域的势函数为

$$g_c(p) = P_p(T \ge c) = \sum_{k=c}^{n} \binom{n}{k} p^k (1-p)^{n-k} ,$$

它关于 p 单调增，关于 c 单调降。固定 c 时犯第一类错误的概率上限为

$$\max_{p \le p_0} g_c(p) = g_c(p_0) 。$$

若取显著性水平为 a ，则应取 c，使得 $g_c(p_0) \le a$ ，即 c 应满足

$$\sum_{k=c}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} \le a 。$$

而 c 越大，犯第二类错误的概率越大，故应取使上式满足的最小的 c。这样就得到了水平为 a 的拒绝域 $\{\underset{\sim}{x} : T \ge c\}$。

例 9.4 某厂对产品次品率的要求是不超过 5%。现从一天生产的产品中随机抽取 20 个检验，发现有 2 个不合格。请问这天的次品率是否达到要求?

解： 设这天的次品率为 p。现欲通过一个容量为 20 的样本来检验

$$H_0 : p \le 5\% \quad v.s. \quad H_1 : p > 5\% 。$$

取水平 $a = 0.05$。若以 $\{\underset{\sim}{x} : T \ge c\}$ 为拒绝域，则 c 是满足下式的最小整数

$$\sum_{k=c}^{20} \binom{20}{k} (5\%)^k (1-5\%)^{n-k} \le a 。$$

取 c ＝0，1，2，4，5 代入上式，发现 $c = 4$ 时上式刚满足。所以拒绝域为 $\{\underset{\sim}{x} : T \ge 4\}$。而 T 的观测值为 2，所以接受原假设。

2. 似然比检验法

似然比检验法，是构造检验的一种较一般的方法。

设样本 $\underset{\sim}{X}=(X_1,\mathbf{L},X_n)$ 的联合密度函数（或分布列）为 $p(\underset{\sim}{x};q)$，$q\in\Theta$。考虑假设检验问题

$$H_0:q\in\Theta_0 \quad v.s. \quad H_1:q\in\Theta_1。$$

显然 $\Theta_0\ \mathbf{I}\ \Theta_1=\varnothing,\ \Theta_0\ \mathbf{U}\ \Theta_1=\Theta$。令

$$l(\underset{\sim}{x})=\frac{\sup\limits_{q\in\Theta_0}p(\underset{\sim}{x};q)}{\sup\limits_{q\in\Theta}p(\underset{\sim}{x};q)}=\frac{p(\underset{\sim}{x};\hat{q}_0)}{p(\underset{\sim}{x};\hat{q}_{MLE})},$$

其中 $\hat{q}_0$ 是 $p(\underset{\sim}{x};q)$ 在参数子空间 Θ_0 上的极大值点，$\hat{q}_{MLE}$ 是 $p(\underset{\sim}{x};q)$ 在参数空间 Θ 上的极大值点，即 q 的极大似然估计值。它们都是样本观测值 $\underset{\sim}{x}$ 的函数，与未知参数无关。所以，$l(\underset{\sim}{X})$ 是一个统计量，称为似然比统计量。由于 $\Theta_0\subset\Theta$，所以 $0\le l(\underset{\sim}{X})\le 1$。

因为 $p(\underset{\sim}{x};q)$ 可看作在给定 $\underset{\sim}{x}$ 时 q 有“多大可能”出现的一种度量。直观地看，H_0 真时，$l(\underset{\sim}{X})$ 较接近于 1，否则较接近于 0。因此可取拒绝域为 $\{\underset{\sim}{x}:l(\underset{\sim}{X})\le c\}$，其中 c $(0<c<1)$ 为检验的临界值。水平取为 a 时，c 应取满足

$$\sup_{q\in\Theta_0}P_q(l(\underset{\sim}{X})\le c)\le a$$

的最大值点。该检验称为水平为 a 的似然比检验。

如果存在统计量 $G(\underset{\sim}{X})$，$l(\underset{\sim}{X})$ 将随之严格上升，那么可以 $\{\underset{\sim}{x}:G(\underset{\sim}{X})\le A\}$ 为拒绝域。

例 9.5 设样本 $X_1,\mathbf{L},X_n\ \ i.i.d.\ N(m,s^2)$，$m\in\mathbf{¡},s^2\in(0.\infty)$ 都未知，欲检验

$$H_0:m=m_0 \quad v.s. \quad H_1:m\ne m_0,$$

其中 m_0 是给定常数。

$\underset{\sim}{X}$ 的联合密度函数是

$$p(\underset{\sim}{x};m,s^2)=(2ps^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2s^2}\sum_{i=1}^{n}(x_i-m)^2\right],$$

这里参数空间 $\Theta=\{(m,s^2):-\infty<m<\infty.\ s^2>0\}$，$\Theta_0=\{(m_0,s^2):s^2>0\}$。在 Θ 上，m,s^2 的 MLE 分别是：$\hat{m}=\bar{X},\quad \hat{s}^2=\sum\limits_{i=1}^{n}(X_i-\bar{X})/n$；在 Θ_0 上，s^2 的 MLE 为

$\hat{s}_0^2 = \sum_{i=1}^{n}(X_i - m_0)^2 / n$。因而

$$l(\underset{\sim}{X}) = \left[\sum_{i=1}^{n}(X_i - \bar{X})^2 / \sum_{i=1}^{n}(X_i - m_0)^2\right]^{n/2} = \left[\frac{1}{1+T^2/(n-1)}\right]^{n/2},$$

其中 $T = \sqrt{n}(\bar{X} - m_0)/S$。由于 $l(\underset{\sim}{X})$ 关于 $|T|$ 严格单调降，因此，似然比检验的拒绝域等价于 $\{\underset{\sim}{x} : |T| \geq c\}$。再根据显著性水平确定 $c = t_{1-a/2}(n-1)$。

3. 根据检验统计量的渐近分布构造检验法

Wilks 定理：设简单随机样本 $X_1, \mathbf{L}, X_n$ 取自具有密度函数（或分布列）的总体，总体分布族为 $\mathrm{F} = \{p(x;q) : q \in \Theta\}$。又参数空间 $\Theta = \Theta_0 \mathbf{U} \Theta_1$，$\Theta_0 \mathbf{I} \Theta_1 = \varnothing$ 为有限维欧氏空间的子集。对假设检验问题 $H_0 : q \in \Theta_0$ *v.s.* $H_1 : q \in \Theta_1$，在分布族满足一定正则性条件下，当原假设为真时，

$$-2\log l(\underset{\sim}{X}) \to_D c^2(k), \quad (k = \dim(\Theta) - \dim(\Theta_0)),$$

其中，$l(\underset{\sim}{X})$ 为似然比检验统计量。

三、关于正态总体参数的假设检验

对于正态总体，常用的对有关参数的假设检验方法有以下六类：

1. 方差已知时，总体均值的检验—— Z 检验（U 检验）
2. 方差未知时，总体均值的检验——t 检验
3. 均值未知时，总体方差的检验——c^2 检验
4. 两总体的方差比较——F 检验
5. 两总体的均值比较——t 检验
6. 成对数据的均值比较

1. 方差已知时，总体均值的检验。$X_1, \mathbf{L}, X_n$ *i.i.d.* $N(m, s^2)$，s^2 已知。

H_0	H_1	检验统计量	水平为 a 的拒绝域
$m \geq m_0$ （或 $m = m_0$）	$m < m_0$	$Z = \dfrac{\bar{X} - m_0}{s/\sqrt{n}}$	$\{(x_1, \mathbf{L}, x_n) : Z \leq z_a\}$
$m \leq m_0$ （或 $m = m_0$）	$m > m_0$		$\{(x_1, \mathbf{L}, x_n) : Z \geq z_{1-a}\}$
$m = m_0$	$m \neq m_0$		$\{(x_1, \mathbf{L}, x_n) : \|Z\| \geq z_{1-a/2}\}$

2. 方差未知时，总体均值的检验——t 检验。$X_1, \mathbf{L}, X_n$ $i.i.d.\ N(m, s^2)$，s^2 未知

H_0	H_1	检验统计量	水平为 a 的拒绝域
$m \geq m_0$ (或 $m = m_0$)	$m < m_0$	$T = \dfrac{\overline{X} - m_0}{S/\sqrt{n}}$	$\{(x_1, \mathbf{L}, x_n): T \leq t_a(n-1)\}$
$m \leq m_0$ (或 $m = m_0$)	$m > m_0$		$\{(x_1, \mathbf{L}, x_n): T \geq t_{1-a}(n-1)\}$
$m = m_0$	$m \neq m_0$		$\{(x_1, \mathbf{L}, x_n): \lvert T \rvert \geq t_{1-a/2}(n-1)\}$

3. 均值未知时，总体方差的检验——c^2 检验。$X_1, \mathbf{L}, X_n$ $i.i.d.\ N(m, s^2)$，m 未知

H_0	H_1	检验统计量	水平为 a 的拒绝域
$s^2 \geq s_0^2$(或 $s^2 = s_0^2$)	$s^2 < s_0^2$	$c^2 = \dfrac{(n-1)S^2}{s_0^2}$	$\{(x_1, \mathbf{L}, x_n): c^2 \leq c_a^2(n-1)\}$
$s^2 \leq s_0^2$(或 $s^2 = s_0^2$)	$s^2 > s_0^2$		$\{(x_1, \mathbf{L}, x_n): c^2 \geq c_{1-a}^2(n-1)\}$
$s^2 = s_0^2$	$s^2 \neq s_0^2$		$\{(x_1, \mathbf{L}, x_n): c^2 \leq c_{a/2}^2(n-1)$ or $c^2 \geq c_{1-a/2}^2(n-1)\}$

4. 两总体的方差比较——F 检验 $d = s_1^2 / s_2^2$

H_0	H_1	检验统计量	水平为 a 的拒绝域
$d \geq d_0$ (或=)	$d < d_0$	$F = \dfrac{S_1^2/S_2^2}{d_0}$	$\{(x, y): F \leq F_a(m-1, n-1)\}$
$d \leq d_0$ (或=)	$d > d_0$		$\{(x, y): F \geq F_{1-a}(m-1, n-1)\}$
$d = d_0$	$d \neq d_0$		$\{(x, y): F \leq F_{a/2}(m-1, n-1)$ or $F \geq F_{1-a/2}(m-1, n-1)\}$

5. 两总体的均值比较——t 检验 $q=m_1-m_2$

	H_0	H_1	检验统计量	水平为 a 的拒绝域
s_1^2, s_2^2 已知	$q \geq q_0$ (或=)	$q < q_0$	$Z=\dfrac{\bar{X}-\bar{Y}-q_0}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}}$	$\{(x_1,\mathbf{L},x_n): Z \leq z_a\}$
	$q \leq q_0$ (或=)	$q > q_0$		$\{(x_1,\mathbf{L},x_n): Z \geq z_{1-a}\}$
	$q = q_0$	$q \neq q_0$		$\{(x_1,\mathbf{L},x_n): \lvert Z \rvert \geq z_{1-a/2}\}$
$s_1^2 = s_2^2$ 未知	$q \geq q_0$ (或=)	$q < q_0$	$T=\dfrac{\bar{X}-\bar{Y}-q_0}{S_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$	$\{(\underset{\sim}{x},\underset{\sim}{y}): T \leq t_a(m+n-2)\}$
	$q \leq q_0$ (或=)	$q > q_0$		$\{(\underset{\sim}{x},\underset{\sim}{y}): T \geq t_{1-a}(m+n-2)\}$
	$q = q_0$	$q \neq q_0$		$\{(\underset{\sim}{x},\underset{\sim}{y}): \lvert T \rvert \geq t_{1-a/2}(m+n-2)\}$

s_1^2, s_2^2 均未知

	H_0	H_1	检验统计量	水平近似为 a 的拒绝域
m, n 都充分大	$q \geq q_0$ (或=)	$q < q_0$	$Z=\dfrac{\bar{X}-\bar{Y}-q_0}{\sqrt{\dfrac{S_1^2}{m}+\dfrac{S_2^2}{n}}}$	$\{(x_1,\mathbf{L},x_n): Z \leq z_a\}$
	$q \leq q_0$ (或=)	$q > q_0$		$\{(x_1,\mathbf{L},x_n): Z \geq z_{1-a}\}$
	$q = q_0$	$q \neq q_0$		$\{(x_1,\mathbf{L},x_n): \lvert Z \rvert \geq z_{1-a/2}\}$
m, n 不都充分大	$q \geq q_0$ (或=)	$q < q_0$	$T=\dfrac{\bar{X}-\bar{Y}-q_0}{\sqrt{\dfrac{S_1^2}{m}+\dfrac{S_2^2}{n}}}$	$\{(\underset{\sim}{x},\underset{\sim}{y}): T \leq t_a(l)\}$
	$q \leq q_0$ (或=)	$q > q_0$	$l=\dfrac{\left(\dfrac{S_1^2}{m}+\dfrac{S_2^2}{n}\right)^2}{\dfrac{S_1^4}{m^2(m-1)}+\dfrac{S_2^4}{n^2(n-1)}}$	$\{(\underset{\sim}{x},\underset{\sim}{y}): T \geq t_{1-a}(l)\}$
	$q = q_0$	$q \neq q_0$		$\{(\underset{\sim}{x},\underset{\sim}{y}): \lvert T \rvert \geq t_{1-a/2}(l)\}$

检验的 p 值

在实际应用中，检验的显著性水平视具体问题而定。为设计软件方便，人们提出 p-值的概念。

一般地，若对假设检验问题

$$H_0: q \in \Theta_0 \quad v.s. \quad H_1: q \in \Theta_1,$$

水平为 a 的拒绝域为 W_a，它满足

$$W_a \subset W_{a'}, \quad \text{if } a < a',$$

对于样本观测值 $\underset{\sim}{x}$，称

$$p = \inf\{a : \underset{\sim}{x} \in W_a\}$$

为观测值 $\underset{\sim}{x}$ 的 ***p*-值 (p-value)**。

p-值的直观含义是：在样本观测值给定的情况下，可以拒绝原假设的最小显著性水平。根据 p-值很容易实现水平为 a 的检验：

当 $p \le a$ 时，拒绝原假设；否则接受原假设。

假设检验与置信区间之间的关系

假设检验与置信区间之间有密切关系。由单参数假设检验问题的水平为 a 的检验，往往可以得到该参数的置信水平为 $1-a$ 的置信区间。反之亦然。

例如，$X_1, \mathbf{L}, X_n$ *i.i.d.* $N(m, s^2)$，m, s^2 均未知。检验 $H_0: m = m_0$ *v.s.* $H_1: m \neq m_0$ 的水平为 a 的拒绝域为

$$W = \{\underset{\sim}{x} : \left| \frac{\bar{X} - m_0}{S/\sqrt{n}} \right| \ge t_{1-a/2}(n-1)\},$$

接受域为

$$\begin{aligned} \bar{W} &= \{\underset{\sim}{x} : \left| \frac{\bar{X} - m_0}{S/\sqrt{n}} \right| < t_{1-a/2}(n-1)\} \\ &= \{\underset{\sim}{x} : \bar{X} - t_{1-a/2}(n-1)\frac{S}{\sqrt{n}} < m_0 < \bar{X} + t_{1-a/2}(n-1)\frac{S}{\sqrt{n}}\}. \end{aligned}$$

若把上述接受域中的 m_0 改为 m，那么由此接受域正好得到 m 的置信水平为 $1-a$ 的置信区间。反之，由 m 的置信水平为 $1-a$ 的置信区间，亦可得到检验问题 $H_0: m = m_0$ *v.s.* $H_1: m \neq m_0$ 的水平为 a 的检验。

一般地，若假设检验问题 $H_0: q = q_0 \ v.s. H_1: q \neq q_0$ 的水平为 a 的检验的接受域若能等价地写成 $\bar{W} = \{\underset{\sim}{x}: \hat{q}_L(\underset{\sim}{x}) < q_0 < \hat{q}_U(\underset{\sim}{x})\}$ 的形式，那么 $(\hat{q}_L(\underset{\sim}{X}), \hat{q}_U(\underset{\sim}{X}))$ 就是q 的置信水平为 $1-a$ 的置信区间。反之，若 $(\hat{q}_L(\underset{\sim}{X}), \hat{q}_U(\underset{\sim}{X}))$ 是q 的置信水平为 $1-a$ 的置信区间，那么，对于$H_0: q = q_0$ $v.s.$ $H_1: q \neq q_0$，下述规则就是一个水平为 a 的检验：

当 $q_0 \in (\hat{q}_L(\underset{\sim}{X}), \hat{q}_U(\underset{\sim}{X}))$ 时接受原假设；否则拒绝原假设。

四、分布拟合检验——c^2 方法

前面讨论的一些假设检验问题主要是对总体未知参数的检验，是参数型的。有时候需要对总体分布的形式作检验，这类检验问题统称为分布拟合检验。c^2 法是一类常用的分布拟合检验法，也是分析属性数据的常用方法。

1．不含未知参数的分布拟合检验

设总体可分为 $A_1, A_2, \mathbf{L}, A_r$ r 类。从中抽取一个容量为 n 的简单随机样本，记各类出现的观测频数分别为 $n_1, n_2, \mathbf{L}, n_r$，$\sum_{j=1}^{r} n_j = n$。欲检验：

$$H_0: P(A_j) = p_{0j},\ j = 1, \mathbf{L}, r \quad v.s. \quad H_1: H_0 \text{不真},$$

其中 $p_{0j} \geq 0,\ j = 1, \mathbf{L}, r$ 是给定的常数，且满足 $\sum_{j=1}^{r} p_{0j} = 1$。

若 H_0 成立，那么当 n 较大时，对每一类A_j，样本频率 n_i/n 应与 p_{0j} 比较接近。K. Pearson 提出用下述统计量构造检验：

$$c^2 = \sum_{j=1}^{r} \frac{(n_j - np_{0j})^2}{np_{0j}} = \sum_{j=1}^{r} \frac{n_j^2}{np_{0j}} - n,$$

并证明了H_0 成立时若 n 足够大，则 $c^2 \ \& \ c^2(r-1)$。因而，水平为 a 的拒绝域为

$$\{c^2 \geq c_{1-a}^2(r-1)\}。$$

2．含未知参数的分布拟合检验

上述检验问题中，常有诸 $p_{0j},\ j = 1, \mathbf{L}, r$ 依赖于 k (k<r) 个未知参数 $q_1, \mathbf{L}, q_k$ 的情形。即欲检验的问题是：$H_0: P(A_j) = p_j(q_1, \mathbf{L}, q_k),\ j = 1, \mathbf{L}, r$。此时须对 c^2 法作修改。c^2 统

计量修改为

$$c^2 = \sum_{j=1}^{r} \frac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j},$$

其中 $\hat{p}_j = p_j(\hat{q}_1, \mathbf{L}, \hat{q}_k)$, $(\hat{q}_1, \mathbf{L}, \hat{q}_k)$ 是 MLE。R. A. Fisher 在 1924 年证明了，在一定条件下，当 H_0 成立时若 n 足够大，则 $c^2 \sim c^2(r-k-1)$。于是水平为 a 的拒绝域为

$$\{c^2 \geq c^2_{1-a}(r-k-1)\}。$$

例 9.6 某种动物的后代按体格的属性分为三类，各类的数目是：10， 53， 46。按照某种遗传模型，其频率之比应为 $p^2 : 2p(1-p) : (1-p)^2$。问数据与模型是否相符？（$a = 0.05$）

解：

记后代的体格属性为 X。考察数据与模型是否相符，就是要检验：

$$H_0 : P(X=1) = p^2,\ P(X=2) = 2p(1-p),\ P(X=3) = (1-p)^2,$$

其中 p 是未知参数，$p \in [0,1]$；H_1 是 H_0 的对立面。可采用 c^2 法检验。

首先，求 H_0 真时 p 的 MLE。若样本 $X_1, \mathbf{L}, X_n$ $i.i.d.$ X，则 H_0 真时其联合分布为

$$\begin{aligned} P(X_1 = x_1, \mathbf{L}, X_n = x_n) &= \prod_{i=1}^{n} \left\{ (p^2)^{I(x_i=1)} [2p(1-p)]^{I(x_i=2)} [(1-p)^2]^{I(x_i=3)} \right\} \\ &= 2^{n_2} p^{2n_1+n_2} (1-p)^{2n_3+n_2} \end{aligned}$$

其中 $n_1 = \sum_{i=1}^{n} I(x_i = 1),\ n_2 = \sum_{i=1}^{n} I(x_i = 2),\ n_3 = n - n_1 - n_2$， $x_i = 1, 2, or\ 3,\ i = 1, \mathbf{L}, n$。

由样本数据知，$n = 109,\ n_1 = 10,\ n_2 = 53,\ n_3 = 46$。因此该样本的似然函数为

$$\begin{aligned} L(p) &= (p^2)^{10} [2p(1-p)]^{53} [(1-p)^2]^{46} \\ &= 2^{53} p^{73} (1-p)^{145}, \quad p \in [0,1] \end{aligned}$$

所以，p 的 MLE 为 $\hat{p} = \frac{73}{73+145} = 0.3349$。

其次，列表计算各类的理论频数与 c^2 统计量。

类别	实际频数 n_i	理论频数 $n\hat{p}_i$	$(n_i - n\hat{p}_i)^2 / (n\hat{p}_i)$
1	10	12.22248	0.4041
2	53	48.55505	0.4069
3	46	48.22248	0.1024
合计	109	109	0.9135

水平 a 的检验规则为：当 $c^2 \geq c^2_{1-a}(3-1-1)$ 时拒绝原假设。$a=0.05$ 时，$c^2_{0.95}(1)=3.841>c^2=0.9135$，因而接受原假设。认为数据与模型相符。

3. 独立性与齐一性检验

设总体中的每一个个体都可按属性 A、B 分类。A 有 r 个类别，B 有 c 个类别。从总体中抽取一个容量为 n 的随机样本，将其中同时具有属性 A_iB_j 的个体的频数记为 n_{ij}，得到如下列联表。

	B_1	B_2	…	B_c	合计
A_1	n_{11}	n_{12}	…	n_{1c}	$n_{1.}$
…	…	…	…	…	…
A_r	n_{r1}	n_{r2}	…	n_{rc}	$n_{r.}$
合计	$n_{.1}$	$n_{.2}$	…	$n_{.c}$	n

独立性检验：$H_0: p_{ij}=p_{i.}p_{.j}, i=1,\mathbf{L}\, r$。$c^2$ 统计量为

$$c^2=\sum_{i=1}^{c}\sum_{j=1}^{r}\frac{(n_{ij}-n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n},$$

自由度 $df=(r-1)(c-1)$，于是水平为 a 的拒绝域为

$$\{c^2 \geq c^2_{1-a}(df)\}。$$

齐一性检验不同在于：抽样方法、原假设。计算、规则都与独立性检验相同。

例 9.7 一项致癌性研究旨在考察一种准备试用于人身上的药物是否有可能导致肿瘤。为此，总共用 300 只（150 只雄性，150 只雌性）老鼠进行为期 6 个月的试验。开始时，随机地将 100 只（50 只雄性，50 只雌性）分配到控制组，100 只分配到低剂量药物组，剩下的 100 只（50 只雄性，50 只雌性）分配到高剂量组。在 6 个月期间，每天给控制组注射一次惰性溶液，而给药物组注射一次掺有药物的溶液。样本数据如下。

肿瘤数目 / 组别	一个或一个以上	无
控制组	10	90
低剂量组	14	86
高剂量组	19	81

问：在显著性水平 $a=0.05$ 时，该药物是否存在与肿瘤有关的药物问题？也就是说，随着药物剂量的增加，患有肿瘤的老鼠的比例是否增加？

解：欲检验 H_0：各剂量组内患肿瘤的老鼠的比例相同。H_1：患肿瘤老鼠的比例不同。

采用 c^2 法检验，水平 a 的检验规则为：当 $c^2 \geq c^2_{1-a}(df)$ 时拒绝原假设。计算列表如下。H_0 真时，各类的期望频数列于下表每格中的第二行。每格中的第三行为：

$$(\text{实际频数}-\text{期望频数})^2 / \text{期望频数}$$

c^2 统计量的值为 6 个格子中的第三行数据之和 3.312，$df=(3-1)(2-1)=2$。$c^2_{0.95}(2)=5.99$。因此，接受 H_0，认为没有充分证据说明患肿瘤的老鼠比例随药物剂量增加而增加。

肿瘤数目 / 组别	一个或一个以上	无	合计
控制组	10	90	100
	14.33	85.67	
	1.310	0.219	
低剂量组	14	86	100
	14.33	85.67	
	0.008	0.001	
高剂量组	19	81	100
	14.33	85.67	
	1.519	0.254	
合计	43	257	300