# 概率论与数理统计

向　稀

2024.09

## 目　录

# 1 随机事件与概率

## 1.1 随机事件及其运算

### 1.1.1 随机试验与随机事件

非确定性现象也称随机现象. 随机现象可以通过随机试验来研究. 所谓随机试验是指对随机现象的一次观测. 把随机试验记为 $E$. 随机试验具有如下特点: 1. 在相同的条件下试验可重复进行; 2. 每次试验具有多种结果, 但试验之前可知道试验的所有可能结果; 3. 每次试验会出现这些可能结果之一, 但试验前不能确定哪一个结果会出现.

在随机试验 $E$ 中, 把所有可能结果的集合称为样本空间, 记为 $\Omega$ . 样本空间的元素, 即 $E$ 的每个可能结果称为**基本事件**或称**样本点**, 记为 $e$ .

随机事件是可能发生也可能不发生的事件. 从随机试验的角度, 可以把随机事件定义为样本空间 $\Omega$ 的子集合. 随机事件常记为 $A$,$B$,$C$ 等. 随机事件可以简称为事件.

随机事件的发生: 该随机事件所包含的某个样本点在随机试验 $E$ 中出现, 则称该随机事件发生.

两个特殊的事件:

必然事件: 样本空间 $\Omega$ . 因为每次试验必有 $\Omega$ 中的样本点出现, 因此 $\Omega$ 必然发生.

不可能事件: 空集 $\Phi$ . 因为 $\Phi$ 不含 $\Omega$ 中任何样本点, 每次试验都没有 $\Omega$ 中的样本点出现, 因此 $\Phi$ 不可能发生.

### 1.1.2 事件间的关系及运算

设 $\Omega$ 是给定的样本空间, 此样本空间上的随机事件为 $A$,$B$,$C$,$A_k(k=1,2,...)$.

由于随机事件定义为样本空间 $\Omega$ 的子集合, 因此事件的关系与运算和集舍的关系与运算是一致的

常用 Venn 图来表示事件的关系与运算. 即用矩形表示样本空间 $\Omega$ , 用 $\Omega$ 内的几何图形表示随机事件.

(一) 事件间的关系

1. 包含关系

如果事件 $A$ 发生必导致事件 $B$ 发生, 则称事件 $B$ 包含事件 $A$ , 记 $B\supset A$.

2. 互不相容关系

若事件 $A$ 与 $B$ 不可能同时发生, 则称 $A$,$B$ 互不相容 (或互斥). 此时,$A$ 与 $B$ 没有共同的样本点.

3. 相等关系

若 $B\supset A$, 且 $A\supset B$ , 则称 $A$ 与 $B$ 相等, 记为 $A=B$. $A$,$B$ 相等时, 两事件的样本点完全相同.

(二) 事件的运算

1. 事件的并

事件 $A$ 与 $B$ 至少发生一个所构成的事件称为 $A$ 与 $B$ 的并, 记为 $A\cup B$ . $A\cup B$ 的样本点由 $A$ 与 $B$ 的样本点合并而成.

推广: 事件 $A_1,A_2,...,A_n$ 至少发生一个的事件称为 $A_1,A_2,...,A_n$ 的并, 记为 $A_1\cup A_2\cup...\cup A_n$ 或 $\bigcup\limits_{i=1}^{n} A$.

2. 事件的交

事件 $A$ 与 $B$ 同时发生的事件称为 $A$ 与 $B$ 的交, 记为 $A \cap B$ 或 $AB.A \cap B$ 的样本点由 $A$ 与 $B$ 共同的样本点组成.

推广: 事件 $A_1, A_2, ..., A_n$ 同时发生的事件称为 $A_1, A_2, ..., A_n$ 的交, 记为 $A_1 \cap A_2 \cap ... \cap A_n$ 或 $\bigcap\limits_{i=1}^{n} A$.

3. 事件的差

事件 $A$ 发生而 $B$ 不发生的事件称为 $A$ 与 $B$ 的差, 记为 $A - B$. $A - B$ 的样本点由 $A$ 中除去 $B$ 的样本点组成.

4. 对立事件

$A$ 不发生的事件称为 $A$ 的对立事件, 记为 $\bar{A}$. $\bar{A}$ 的样本点由 $\Omega$ 中除 $A$ 以外的样本点组成.

注意如下等式:

(1) $A = (A - B) \cup AB$.

(2) $A \cup B = A \cup (B - A) = (A - B) \cup AB \cup (B - A)$, 第二个等式表示 $A \cup B$ 可以分为互不相容的三个部分.

(3) $A - B = A - AB = AB$, 第二个等式表示事件的差可以用事件的交表示.

(4) $\bar{A} = \Omega - A, A \cup A = \Omega$.

可以证明事件的运算满足以下的运算律:

交换律:$A \cup B = B \cup A, A \cap B = B \cap A$;

结合律:$(A \cup B) \cup C = A \cup (B \cup C), (A \cap B) \cap C = A \cap (B \cap C)$;

分配律:$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;

德摩根 (De Morgan) 定律:$A \bar{\cup} B = \bar{A} \cap \bar{B}, A \bar{\cap} B = \bar{A} \cup \bar{B}$.

可以推广到 $n$ 个的情形:

$$A_1 \cup A_2 \bar{\cup} ... \cup A_n = \bar{A}_1 \cap \bar{A}_2 \cap ... \cap \bar{A}_n,$$
$$A_1 \cap A_2 \bar{\cap} ... \cap A_n = \bar{A}_1 \cup \bar{A}_2 \cup ... \cup \bar{A}_n.$$

## 1.2 事件的概率及性质

### 1.2.1 频率与概率

对于一个随机事件, 希望知道其发生可能性的大小. 我们可以通过随机试验 $E$ 来了解这一可能性. 设随机事件 $A$ 在 $n$ 次重复试验中共出现 $n_A$ 次, 将其比值 $n_A/n$ 定义为 $A$ 发生的频率, 记为 $f_n(A)$. 即

$$f_n(A) = \frac{n_A}{n}.$$

直观上看, 频率可以在一定程度上反映事件发生可能性的大小.

事件 $A$ 频率具有如下性质:

(1) $0 \leq f_N(A) \leq 1$;

(2) $f_n(\Omega) = 1$; (3) 若 $A_1, A_2, ..., A_n$ 互不相容, 则

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

**频率的稳定性** 在 $n$ 次试验中, 随着试验次数 $n$ 的增加, 事件的频率将在某个数值 $p$ 附近稳定地摆动, 一般来说,$n$ 越大, 摆动的幅度越小. 此数值可以定义为 $A$ 发生的概率, 称为**统计概率**.

### 1.2.2 概率的定义及性质

1933 年, 苏联数学家柯尔莫哥洛夫提出了概率的公理化定义, 即通过规定概率应具有的基本性质来定义概率. 柯尔莫哥洛夫提出的公理很简洁, 但在此基础上建立起了概率论的理论体系.

**定义 1.1** 设 E 是随机试验,$\Omega$ 为其样本空间. 对于 $E$ 的每一个随机事件 $A$, 对应一个实数 $P(A)$ 如果集合函数 $P(\cdot)$ 满足:

(1) 非负性 $P(A) \geq 0$;

(2) 规范性 $P(\Omega) = 1$;

(3) 可列可加性 若 $A_1, A_2, ..., A_n, ...$ 两两互不相容, 则

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

则称 $P(A)$ 为事件 $A$ 的概率.

利用概率的公理化定义, 可以得到概率的一些性质.

**性质 1** $P(\Phi) = 0$.

**证明** $\Omega = \Omega \cup \Phi \cup \Phi \cup \Phi ...$, 由可列可加性

$$P(\Omega) = P(\Omega) + P(\Phi) + P(\Phi) + P(\Phi) + ...,$$

从而有 $P(\Phi) = 0$.

**性质 2** 若 $A_1, A_2, ..., A_n$ 为 $n$ 个两两互不相容事件, 则有

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

**证明** 在可列可加性中, 取 $A_{n+1} = A_{n+2} = ... = \Phi$, 因为 $P(\Phi) = 0$, 得证.

**性质 3** 对任意两事件 $A, B$ 有

$$P(A - B) = P(A) - P(AB).$$

**证明** 因为 $A = (A - B) \cup AB$. 而 $A - B$ 与 $AB$ 互不相容, 由性质 2,

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

从而

$$P(A - B) = P(A) - P(AB).$$

同理 $P(B - A) = P(B) - P(AB)$.

当 $A \supset B$ 时,$P(AB) = P(B)$. 从而 $P(A - B) = P(A) - P(B)$ 且 $P(A) \geq P(B)$.

**性质 4** 对任意两事件 $A, B$ 有**加法定理**

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

**证明** 因为 $A \cup B = A \cup (B - A)$. $A$ 与 $B - A$ 互不相容, 再由性质 3, 得到

$$P(A \cup B) = P(A) + P(B - A) = P(A) + P(B) - P(AB).$$

利用归纳法, 可推广到 $n$ 个事件的加法定理

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i A_j) + \sum_{1 \le i < j < k \le n} P(A_i A_j A_k) + ... + (-1)^{n-1} P(A_1 A_2 ... A_n).$$

**性质 5** 对任意事件 $A, P(\bar{A}) = 1 - P(A)$.

**证明** $A \cup \bar{A} = \Omega, A \cap \bar{A} = \Phi$, 由性质 2 得

$$1 = P(\Omega) = P(A \cup \bar{A} = P(A) + P(\bar{A})).$$

从而 $P(\bar{A}) = 1 - P(A)$.

以上概率性质中的公式, 也可在 Venn 图中得到反映. 只要定义 $\Omega$ 的面积为 1, 某事件在 Venn 图的面积对应其概率.

## 1.3 等可能概型 (古典概型)

有一类随机试验, 具有如下特点:

(1) 样本空间 $\Omega$ 的样本点个数为有限个;

(2) 样本空间中的每个基本事件 (样本点) 发生的可能性相同.

把这类随机试验称为等可能概型, 由于它是概率论发展初期的主要研究对象, 又称为**古典概型**.

设随机试验 $E$ 为等可能概型, 若其样本空间. 含有 $n$ 个样本点, 而事件 $A$ 含有 $m$ 个样本点, 则事件 $A$ 的概率为

$$P(A) = \frac{m}{n} = \frac{A\text{包含的样本点数}}{\text{样本点总数}}.$$

这样的定义是合理的. 设 $\Omega = \{e_1, e_2, ..., e_n\}$, 由于等可能,$P(e_i) = 1/n$. 设 $A$ 含有 $m$ 个基本事件,$A = \{e_{i1}, e_{i2}, ..., e_{im}\}$, 则

$$P(A) = P(e_{i1}) + P(e_{i2}) + ... + P(e_{im}) = \frac{1}{n} + \frac{1}{n} + ... + \frac{1}{n} = \frac{m}{n}.$$

**例 1.1** 袋中有 $a$ 只白球,$b$ 只红球. 从中将球依次取出, 问第 $k$ 次取出的球是红球的概率是多少?

**解** 样本点总数为 $a + b$ 只球依次排列的所有种数 $(a + b)!$. 设 $A$: 第 $k$ 个为红球, 则第 $k$ 个红球有 $b$ 种选择, 其余 $a + b - 1$ 个位置共有 $(a + b - 1)!$ 种排法. 因此,$A$ 包含样本点数为 $b(a + b - 1)!$. 得到

$$P(A) = \frac{b(a + b - 1)!}{(a + b)!} = \frac{b}{a + b}.$$

注意此结果与 $k$ 无关. 由此引申出" 抽签原理". 即袋中有 $a$ 只白球,$b$ 只红球. 若干人依次从中各取一球, 取后不放回, 则任一人取得红球的概率相等, 都是 $\frac{b}{a+b}$.

**例 1.2** $N$ 件产品, 含 $M$ 件次品, 其余为正品. 现从中任意取出 $n$ 件, 按不放回抽样和有放回抽样两种情况, 分别求其中恰有 $k(k \leq M)$ 件次品的概率.

**解** **不放回抽样**, 每次取出不放回, 继续抽下一件. 样本点总数为 $N$ 件产品取出 $n$ 件的所有取法 $C_N^n$. 恰有 $k$ 件次品, 则同时取 $n-k$ 件正品, 取次品有 $C_M^k$ 种取法, 取正品有 $C_{N-M}^{n-k}$ 种取法. 因此所求概率为

$$p = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}.$$

**有放回抽样**, 每次取出放回, 再任取下一件. 样本点总数为从 $N$ 件产品中有放回地抽取 $n$ 件的排列种数 $N^n$. 恰有 $k$ 件次品的取法, 取出的 $n$ 个排成一列, 在 $n$ 个位置选 $k$ 个位置放次品 $C_n^k$ 种, 每个次品有 $M$ 种取法, 取次品共 $C_n^k M^k$ 种; 取 $n-k$ 个正品,$(N-M)^{n-k}$ 种. 因此所求概率为

$$p = \frac{C_n^k M^k (N-M)^{n-k}}{N^n} = C_n^k \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{n-k}.$$

注意, 第二个等号后的写法正好对应后面所讲的独立重复试验概型.

**例 1.3** 将 $n$ 只小球随机放入 $N$ 个盒子 $(N \geq n)$ 时, 求以下概率: (1) 事件 $A$: 恰有 $n$ 个盒子每盒 1 球; (2) 事件 $B$: 某指定的 $n$ 个盒子中各有 1 球.

**解** 样本点总数:$n$ 只球放入 $N$ 个盒子, 共有 $N \times N \times ... \times N = N^n$ 种放法.

(1)$n$ 个盒子是可选的, 共有 $C_N^n$ 种.$n$ 个球放入 $n$ 个盒子有 $n!$ 种, 因此 $A$ 含有 $C_N^n n!$ 个样本点, 则

$$P(A) = \frac{C_N^n \cdot n!}{N^n}$$

(2) 盒子指定, 只需放入 $n$ 个球, 有 $n!$ 种放法.$B$ 含 $n!$ 个样本点, 因此

$$P(B) = \frac{n!}{N^n}.$$

本例的概率模型有很多应用, 如统计物理中的麦克斯韦-玻尔兹曼统计, 玻色-爱因斯坦统计. 以下的**生日问题**同样可以用这一模型解决.

**例 1.4** 某班级有 $n$ 个同学, 求至少两人生日相同的概率. 假设每人的生日在 365 天中任一天等可能. **解** 设 $A$: 至少 2 人生日相同.$A$ 是一个复杂的事件. 考虑 $A$ 的对立事件:$n$ 个同学生日都不同. 这相当于将 $n$ 球放入 $N = 365$ 个盒子, 使得恰有 $n$ 个盒子每盒 1 球. 按例 1. 3(1) 的情形, 此概率为

$$P(\bar{A}) = \frac{C_N^n \cdot n!}{N_n}.$$

因此

$$P(A) = 1 - \frac{C_N^n \cdot n!}{N_n}.$$

可以求得,$n = 23$ 时,$P(A) = 0.507$. 已经超过一半;$n = 50$ 时,$P(A) = 0.97$. 这个概率比直观感觉要大.

**例 1.5** 设有 $k$ 个盒子, 每个盒子装有号码 1 到 $n$ 的 $n$ 个球. 现从每个盒子中任取一球. 求取得的 $k$

个球中最大号码为 $m$ 的概率.

**解**

$$P(A)=\frac{m^k-(m-1)^k}{n^k}.$$

### 1.4 几何概率

古典概型中要求样本空间中的样本点总数有限. 而很多实际问题中, 样本空间中的样本点无限, 但各样本点仍具有等可能的特点. 这种情况下, 可以用几何方法计算概率.

**定义 1.2** 若随机试验的样本空间 $\Omega$ 对应一个度量有限的几何区域 $S$, 且所有的基本事件与 $S$ 内的点一一对应, 则任一随机事件 $A$ 对应 $\Omega$ 中的某一子区域 $D$. 若事件 $A$ 的概率只和 $A$ 对应的区域 $D$ 的度量成正比, 与 $D$ 的形状及 $D$ 在 $S$ 中的位置无关.$A$ 发生的概率定义为

$$P(A)=\frac{m(A)}{m(\Omega)}=\frac{A\text{对应区域}D\text{的度量}}{\Omega\text{对应区域}S\text{的度量}}.$$

这样计算的概率称为几何概率.

**例 1.5(会面问题)** 甲乙两人约定 8 点到 9 点间随机到达某地会面, 两人约定先到者等候另一人 15 分钟, 过时则离去. 求两人能会面的概率.

**解** 以 $x,y$ 表示甲乙的到达时刻, 则样本空间 $\Omega=\{(x,y)\ |\ 0\le x\le 60, 0\le y\le 60\}$ (分钟), 为边长 60 的正方形. 设 $A$ 为两人能见面, 则 $A=\{(x,y)\ |\ |\ x-y\ |\le 15\}$.$A$ 为图 1 中的阴影部分.

$$P(A)=\frac{m(A)}{m(\Omega)}=\frac{60^2-45^2}{60^2}=\frac{7}{16}.$$

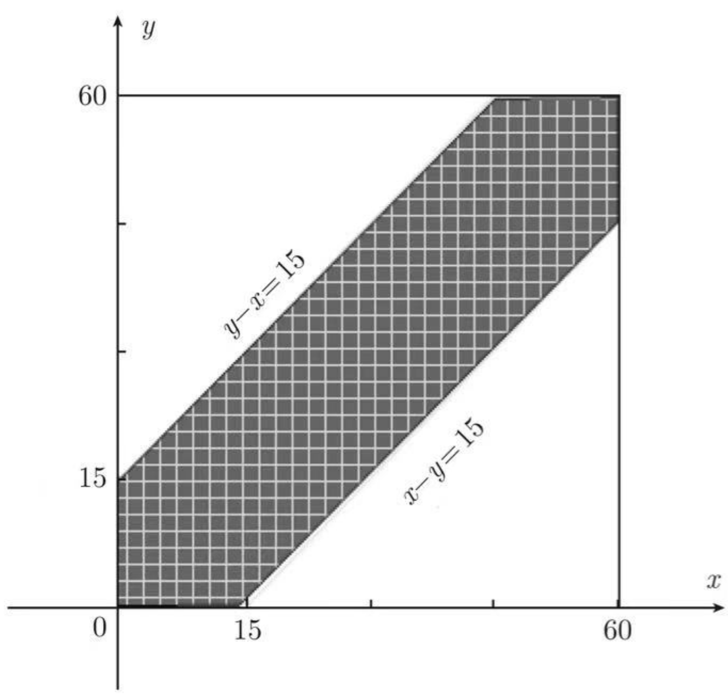


图 1

**例 1.6 (布丰投针问题)** 平面上有等距离的平行线, 平行线间的距离为 $a$. 向此平面投掷一枚长为 $l(l<a)$ 的针, 求针与任一平行线相交的概率.

**解** 设 $M$ 为针的中点,$M$ 点到最近平行线的距离为 $x$, 针与平行线的夹角为 $\theta$, 如图 2 所示. 针的位置可由 $(x,\theta)$ 决定, 可得样本空间

$$\Omega=\left\{(x,\theta)\ \mid\ 0\le x<a/2, 0\le\theta\le\pi\right\}.$$

设 $A$: 针与任一条平行线相交. 其充要条件为

$$x\le\frac{l}{2}\sin\theta.$$

从而 $A=\left\{0\le x\le\frac{l}{2}\sin\theta, 0\le\theta\le\pi\right\}$, 如图 2 所示.

$$P(A)=\frac{m(A)}{m(\Omega)}=\frac{\displaystyle\int_0^\pi\frac{l}{2}\sin\theta\mathrm{d}\theta}{a\pi/2}=\frac{2l}{a\pi}.$$

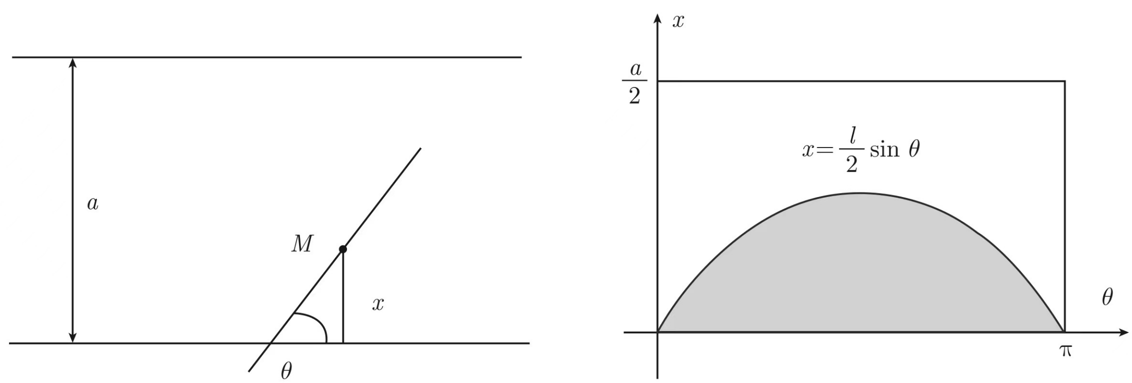


图 2

根据以上结果可以利用随机试验的方法求 $\pi$ 的近似值. 设 $n$ 次投针试验中, 针与平行线相交 $m$ 次. 当 $n$ 较大时, 以频率 $m/n$ 作为 $P(A)$ 的近似值, 代入上式, 得

$$\frac{m}{n}\approx P(A)=\frac{2l}{a\pi}.$$

从而得到 $\pi\approx\dfrac{2ln}{am}$.

## 1.5 条件概率

### 1.5.1 条件概率

除了考虑事件 $A$ 发生的概率, 经常还需要考虑在另一事件 $B$ 发生的情况之下, 事件 $A$ 发生的概率. 这一概率称为条件概率, 即 $B$ 发生的条件下事件 $A$ 发生的概率, 记为 $P(A\mid B)$.

应当怎样求得条件概率 $P(A\mid B)$, 我们作如下分析.

在古典概型之下, 假设样本空间 $\Omega$ 含有 $n$ 个样本点, 事件 $A$ 含有 $i$ 个样本点, 事件 $B$ 含有 $m$ 个样本点, 事件 $AB$ 含有 $k$ 个样本点.

显然

$$P(A)=\frac{i}{n},P(B)=\frac{m}{n},P(AB)=\frac{k}{n}.$$

考虑 $B$ 发生的条件下 $A$ 发生的概率. 既然 $B$ 已经发生, 我们关心的样本点应当缩小至 $B$ 的范围内. 此时 $B$ 可以看成缩小的样本空间, 记为 $\Omega_B.A$ 在 $B$ 内含有的样本点数 $k$( 即 $AB$ 所含的样本点数) 与 $B$ 所含样本点数 $m$ 之比应为所求概率. 因此

$$P(A\mid B)=\frac{k}{m}=\frac{k/n}{m/n}=\frac{P(AB)}{P(B)}.$$

**定义 1.3** 设 $A,B$ 为两个随机事件, 且 $P(B)>0$, 称

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

为在事件 $B$ 发生的条件下, 事件 $A$ 发生的概率.

由条件概率定义, 容易验证条件概率 $P(\cdot\mid B)$ 满足概率定义中的三个条件:

(1) 非负性　对任意事件 $A$,$P(A\mid B)\geq O$;

(2) 规范性　$P(\Omega\mid B)=1$;

(3) 可列可加性　若 $A_1,A_2,...,A_n,...$ 两两互不相容, 则

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

因此, 定义 1.3 中推导的概率性质, 对条件概率仍然适用. 例如, 对任意两事件 $A_1,A_2$ 有

$$P(A_1\cup A_2\mid B)=P(A_1\mid B)+P(A_2\mid B)-P(A_1A_2\mid B).$$
$$P(A_1-A_2\mid B)=P(A_1\mid B)-P(A_1A_2\mid B).$$

### 1.5.2　乘法公式

由条件概率公式可以得到, 当 $P(B)>0$ 时,

$$P(AB)=P(B)P(A\mid B).$$

这称为乘法公式. 对称地还可得到, 当 $P(A)>0$ 时,

$$P(AB)=P(A)P(B\mid A).$$

乘法公式可以推广到多个的情况

设 $A_1,A_2,...,A_n$ 为 $n$ 个随机事件, 当 $P(A_1A_2...A_n)>0$ 时,

$$P(A_1A_2...A_n)=P(A_1)P(A2\mid A_1)P(A3\mid A_1A_2)...P(A_n\mid A_1,A_2...A_{n-1}).$$

利用乘法公式, 可以把复杂的交事件的概率化为简单事件的概率来求解.

**例 1.7** 一串钥匙 $n$ 把, 只有一把能开门, 任取一把开门, 用后分开, 求第 $k$ 次才打开门的概率.

**解** 设 $A_i$ 是第 $i$ 次**不能**打开门的事件.$(i=1,2,...,k)$. 则第 $k$ 次才打开门的事件为 $A_1A_2...A_{k-1}\bar{A}_k$.

$$\begin{aligned}P(A_1A_2...A_{k-1}\bar{A}_k) &= P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2)...P(A_{k-1}\mid A_1,A_2...A_{k-2})P(\bar{A}_k\mid A_1,A_2...A_{k-1})\\ &= \frac{n-1}{n}\cdot\frac{n-2}{n-1}\cdot\frac{n-3}{n-2}\cdot...\cdot\frac{n-(k-1)}{n-(k-2)}\cdot\frac{1}{n-(k-1)}=\frac{1}{n}.\end{aligned}$$

此结果与 $k$ 无关. 这也与抽签原理相符合.

### 1.5.3 全概率公式与贝叶斯公式

一些复杂事件的概率计算往往需要将复杂事件划分为若干个互不相容的简单事件的并, 然后再分别计算简单事件的概率, 根据概率可加性得到结果. 全概率公式就是这一想法的实现.

**定义 1.4** 若 $n$ 个事件 $A_1,A_2,...,A_n$ 满足

(1) $A_1,A_2,...,A_n$ 两两互不相容;

(2) $A_1\cup A_2\cup...\cup A_n=\Omega$.

则称 $A_1,A_2,...,A_n$ 为样本空间 $\Omega$ 的一个划分或完备事件组.

若事件 $A_1,A_2,...,A_n$ 为完备事件组, 则在随机试验的每次试验中,$A_1,A_2,...,A_n$ 必发生一个且仅发生一个.

**定理 1.1(全概率公式)** 设 $A_1,A_2,...,A_n$ 为完备事件组, 则对事件 $B$ 有

$$P(B)=\sum_{i=1}^{n}P(A_kP(B\mid A_k)).$$

**证明** 因为 $A_1\cup A_2\cup...\cup A_n=\Omega$, 可得:$B=\Omega B=\left(\bigcup_{k=1}^{n}A_k\right)B=\bigcup_{k=1}^{n}A_kB$, 如图 3 所示.

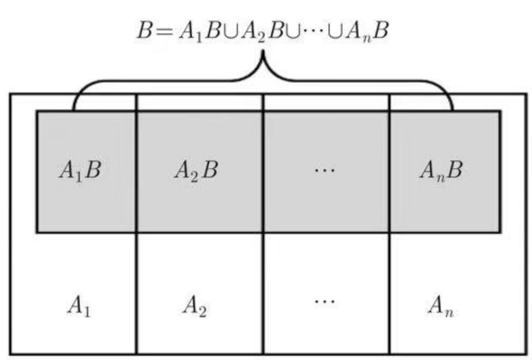


图 3

$$P(B)=P\left(\bigcup_{k=1}^{n}A_kB\right)=\sum_{k=1}^{n}P(A_kB)=\sum_{k=1}^{n}P(A_k)P(B\mid A_k).$$

**例 1.8** 甲袋有 5 个球 (3 红 2 白), 乙袋有 8 个球 (4 红 4 白), 现从甲袋任取 2 球放入乙袋, 再从乙袋任取 1 球, 问取白球的概率?

**解** 从甲袋取出 2 球有 3 种情况, 设 $A_1$: 取出 2 白球;$A_2$: 取出 2 红球;$A_3$: 取出 1 红球和 1 白球. 设 $B$: 最后取白球. 则

$$P(A_1)=\frac{1}{C_5^2}=\frac{1}{10},P(A_2)=\frac{C_3^2}{C_5^2}=\frac{3}{10},P(A_3)=\frac{C_3^1C_2^1}{C_5^2}=\frac{6}{10}.$$

$A_1$ 发生, 则乙袋内 4 红 6 白,$P(B\mid A_1)=\dfrac{6}{10}$, 同理 $P(B\mid A_2)=\dfrac{4}{10}$,$P(B\mid A_3)=\dfrac{5}{10}$.
所以

$$P(B)=\sum_{k=1}^{n}P(A_k)P(B\mid A_k)=\frac{1}{10}\times\frac{6}{10}+\frac{3}{10}\times\frac{4}{10}+\frac{6}{10}\times\frac{5}{10}=\frac{48}{100}.$$

**例 1.9** 设甲袋有 $N-1$ 只白球和 1 只黑球, 乙袋中有 $N$ 只白球. 若每次从甲, 乙两袋中各取一球交换, 记 $A_k$ 为经过 $k$ 次交换后黑球仍在甲袋中, 证明:

$$P(A_{k+1})=\frac{1}{N}+\left(1-\frac{2}{N}\right)P(A_k),\qquad k=1,2,...$$

**证明** $A_k$ 表示 $k$ 次交换后黑球仍在甲袋, 因此,$A_k$ 发生时, 甲袋有 $N-1$ 只白球和 1 只黑球, 此后若 $A_{k+1}$ 发生, 等价于从甲袋取出白球; $\bar{A}_k$ 发生时, 乙袋有 $N-1$ 只白球和 1 只黑球, 此后若 $A_{k+1}$ 发生, 等价于从乙袋取出黑球, 因此

$$P(A_{k+1}\mid A_k)=\frac{N-1}{N},P(A_{k+1}\mid \bar(A_k))=\frac{1}{N}.$$

由全概率公式

$$\begin{aligned}P(A_{k+1})&=P(A_k)P(A_{k+1}\mid A_k)+P(\bar{A}_k)P(A_{k+1|\bar{A}_k})\\&=P(A_k)\frac{N-1}{N}+[1-P(A_k)]\frac{1}{N}\\&=\frac{1}{N}+\left(1-\frac{2}{N}\right)P(A_k).\end{aligned}$$

在全概率公式 $P(B)=\sum_{i=1}^{n}P(A_kP(B\mid A_k))$ 中, 事件 $A_1,A_2,...,A_n$ 可以看成事件 $B$ 发生的原因,$B$ 看成结果,$P(A_k)$ 称为先验概率. 若事件 $B$ 已发生, 此时考虑事件 $A_k$ 发生的概率, 即 $P(A_k\mid B)$. 它反映了结果发生时, 各个原因发生的可能性大小, 称为后验概率.

**定理 1.2** 设 $A_1,A_2,...,A_n$ 为完备事件组, 则对事件 $B$ 有

$$P(A_i\mid B)=\frac{P(A_i)P(B\mid A_i)}{\sum\limits_{k=1}^{n}P(A_k)P(B\mid A_k)}.$$

**证明**

$$P(A_i \mid B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)P(B \mid A_i)}{\sum\limits_{k=1}^{n}P(A_k)P(B \mid A_k)}.$$

**例 1.10** 前面取球的例子: 已知最后从乙袋取出白球, 问从甲袋放入乙袋的 2 球都是红球的概率是多少?

**解** 按前例假设, 已知 $B$ 发生, 求 $A_2$ 发生的概率, 即求 $P(A_2 \mid B)$. 由贝叶斯公式,

$$P(A_2 \mid B)=\frac{P(A_2B)}{P(B)}=\frac{P(A_2)P(B \mid A_2)}{\sum\limits_{k=1}^{3}P(A_k)P(B \mid A_k)}=\frac{\frac{3}{4}\times\frac{4}{10}}{\frac{48}{100}}=\frac{12}{48}.$$

同理, 还可求 $P(A_1 \mid B)=\dfrac{6}{48}, P(A_3 \mid B)=\dfrac{30}{48}$.

由此可见, $P(A_1 \mid B)+P(A_2 \mid B)+P(A_3 \mid B)=1$. 这说明 $B$ 发生时,$A_1,A_2,A_3$ 必发生且仅发生一个, 与 $A_1,A_2,A_3$ 是完备事件组相符合.

**例 1.11** 一批产品中有 96% 是合格品. 现有一种简化的检查方法, 它把合格品判为合格品的概率为 0.98, 而将不合格品误判为合格品的概率为 0.05, 任取一件产品. (1) 求用此方法检查出合格品的概率; (2) 求检查出的合格品确为合格品的概率.

**解** 任取一件产品, 设事件 $A$: 取到合格品,$\bar{A}$ 取到不合格品. 设事件 $B$: 判为合格品.

则 (1) 求 $P(B)$; (2) 求 $P(A \mid B)$.

(1) $P(A)=0.96$,$P(\bar{A})=0.04$,$P(B \mid A)=0.98$, $P(B \mid \bar{A})=0.05$, 由全概率公式,

$$\begin{aligned}P(B)&=P(A)P(B \mid A)+P(\bar{A})P(B \mid \bar{A})\\&=0.96\times0.98+0.04\times0.05\\&=0.942.\end{aligned}$$

(2) 由贝叶斯公式,

$$P(A \mid B)=\frac{P(A)P(B \mid A)}{P(A)P(B \mid A)+P(\bar{A})P(B \mid \bar{A})}=\frac{0.96\times0.98}{0.9428}=0.9979.$$

**例 1.12** 某学生接连参加同一课程的两次考试, 第一次及格的概率为 $p$. 若第一次及格则第二次及格的概率也为 $p$; 若第一次不及格则第二次及格的概率为 $p/2$. (1) 若至少一次及格则他能取得某种资格, 求他取得资格的概率; (2) 若他第二次已经及格, 求他第一次及格的概率.

**解** 设 $A_i$ 为他第 $i$ 次考试及格 $(i=1,2)$ ,$B$ 表示他取得资格. 由题意

$$P(A_1)=P(A_2 \mid A_1)=p, \qquad P(A_2 \mid \bar{A}_1)=p/2.$$

(1)

$$\begin{aligned}P(B) &= P(A_1 \cup P_2) = P(A_1 \cup \bar{A}_1 A_2) = P(A_1) + P(\bar{A}_1 A_2)\\ &= P(A_1) + P(\bar{A}_1)P(A_2 \mid \bar{A}_1) = p + (1-p)\frac{p}{2} = \frac{3}{2}p - \frac{1}{2}p^2.\end{aligned}$$

(2)

$$\begin{aligned}P(A_1 \mid A_2) = \frac{P(A_1A_2)}{P(A_2)} &= \frac{P(A_1)P(A_2 \mid A_1)}{P(A_1)P(A_2 \mid A_1) + P(\bar{A}_1)P(A_2 \mid \bar{A}_1)}\\ &= \frac{p \cdot p}{p \cdot p + \frac{p}{2}(1-p)} = \frac{2p}{p+1}.\end{aligned}$$

## 1.6 独立性

### 1.6.1 独立性定义

两个随机事件 $A$ 和 $B$, 若 $B$ 发生时 $A$ 发生的概率 $P(A \mid B)$ 与 $A$ 发生的概率不相同, 即 $P(A \mid B) \neq P(A)$ , 则可认为 $B$ 的发生对 $A$ 的发生有影响. 若 $P(A \mid B) = P(A)$, 则可认为 $B$ 的发生对 $A$ 的发生没有影响. 这就引出了独立性的概念.

**定义 1.5** 设 $A, B$ 是两个随机事件, 若满足等式

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

则称事件 $A, B$ 相互独立, 简称 $A, B$ 独立.

**定理 1.3** 若随机事件 $A$ 与 $B$ 相互独立, 则 $\bar{A}$ 与 $B$,$A$ 与 $\bar{B}$,$\bar{A}$ 与 $\bar{B}$ 也相互独立.

**证明** 先证 $\bar{A}$ 与 $B$ 独立. 因为 $A$ 与 $B$ 相互独立, $P(AB) = P(A)P(B)$.

$$\begin{aligned}P(\bar{A}B) &= P(B - A) = P(B) - P(AB)\\ &= P(B) - P(A)P(B)\\ &= [1 - P(A)]P(B) = P(\bar{A})P(B).\end{aligned}$$

所以,$A$ 与 $B$ 独立 $\Rightarrow$ $\bar{A}$ 与 $B$ 独立. 再由 $A$ 与 $B$ 的对称性 $\Rightarrow$ $B$ 与 $A$ 独立 $\Rightarrow$ $\bar{B}$ 与 $A$ 独立. 同样由对称性,$A$ 与 $\bar{B}$ 独立 $\Rightarrow$ $\bar{A}$ 与 $\bar{B}$ 独立.

### 1.6.2 多个事件的独立性

设 $A, B, C$ 是三个随机事件, 若

$$
\begin{cases}
P(AB) = P(A)P(B) \\
P(BC) = P(B)P(C) \\
P(AC) = P(A)P(C) \\
P(ABC) = P(A)P(B)P(C)
\end{cases}
$$

则称三事件 $A, B, C$ 相互独立.

一般, 对 $n$ 个事件 $A_1, A_2, ..., A_n$, 如果下列等式成立

$$
\begin{cases}
P(A_iA_j) = P(A_i)P(A_j), 1 \le i < j \le n \\
P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k), 1 \le i < j < k \le n \\
\cdots \\
P(A_{i_1}A_{i_2}...A_{i_m}) = P(A_{i_1})P(A_{i_2})...P(A_{i_m}), 1 \le i_1 < i_2 < \cdots < i_m \le n \\
\cdots \\
P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n)
\end{cases}
$$

则称事件 $A_1, A_2, ..., A_n$ 相互独立.

注意, $n$ 个事件 $A_1, A_2, ..., A_n$ 相互独立, 需要 $C_n^2 + C_n^3 + \cdots + C_n^n = 2^n - n - 1$ 个条件. $n$ 个事件相互独立, 则其中任意 $k(2 \le k \le n)$ 个事件也相互独立.

相互独立能推出两两独立, 反之不然.

**补充:** $A_1, A_2, A_3$ 两两独立, 则有 $P((A_1 \cup A_2)A_3) = P(A_1 \cup A_2)P(A_3)$.

**证明** $P((A_1 \cup A_2)A_3) = P(A_1A_3 \cup A_2A_3) = P(A_1A_3) + P(A_2A_3) = P(A_1)P(A_3) + P(A_2)P(A_3) = (P(A_1) + P(A_2))P(A_3) = P(A_1 \cup A_2)P(A_3)$.

以下不加证明地给出分组独立性定理, 在讨论独立性问题时很有用处.

**定理 1.4** 设事件 $A_1, A_2, ..., A_n$ 相互独立, 将其任意分为没有公共事件的 $k$ 个组, 每个组任意作事件运算, 得到一个新事件, 则这 $k$ 个新事件相互独立.

$n$ 个事件的加法定理很复杂, 但在事件 $A_1, A_2, ..., A_n$ 相互独立情况下, 运算得到简化, 可以表示为

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n).$$

推导如下

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \cdots \cup A_n) &= 1 - P(A_1 \cup A_2 \bar{\cup} \cdots \cup A_n) \\
&= 1 - P(\bar{A}_1\bar{A}_2\cdots\bar{A}_n) \\
&= 1 - P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n).
\end{aligned}
$$

**例 1.13** 设随机试验中某事件 $A$ 发生的概率为 $p$ , 试证明, 无论 $p > 0$ 多么小, 只要不断独立重复地做 $n$ 次试验, $A$ 迟早会发生的概率为 1.

**证明** 设 $A_i$ 为第 $i$ 次试验中 $A$ 发生, 则 $n$ 次试验 $A$ 都不出现的事件为 $\bar{A}_1\bar{A}_2\cdots\bar{A}_n$.

$$
\begin{aligned}
P(\bar{A}_1\bar{A}_2\cdots\bar{A}_n) &= P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n)\\
&= (1-p)(1-p)\cdots(1-p)\\
&= (1-p)^n \to 0.
\end{aligned}
$$

则 $n$ 次试验中 $A$ 至少出现一次的概率

$$P(A_1\cup A_2\cup\cdots\cup A_n) = 1-(1-p)^n \to 1.$$

这一结论也称为**小概率事件原理**.

## 1.7 重复独立试验概型

有一类重要的独立重复试验概型, 具有如下特点:

(1) 每次试验只有两个结果,$A$ 和 $\bar{A}$, 且 $P(A)=p$, $P(\bar{A})=1-p=q$.

(2) 试验进行 $n$ 次, 每次试验的结果相互独立.

这样的试验我们称为 $n$ 重伯努利试验. 例如, 一枚硬币抛 $n$ 次, 每次两个结果: 正面或反面向上. 或者产品检验中的有放回抽样 $n$ 次, 每次两个结果: 正品或次品. 这类试验中, 我们所关心的是, $n$ 次试验中 $A$ 发生 $k$ 次的概率. 一般, 我们有

**定理 1.5** $n$ 重伯努利试验中, $A$ 发生 $k$ 次的概率记为 $P_n(k)$

$$P_n(k) = C_n^k p^k q^{n-k}, k=0,1,...,n; q=1-p.$$

注意,$\displaystyle\sum_{k=0}^{n} P_n(k) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1.$

**例 1.13** 利用独立重复试验概型, 解有放回抽样情形. 即 $N$ 件产品, 含 $M$ 件次品, 其余为正品. 现从中任意取出 $n$ 件, 按有放回抽样情况, 求其中恰有 $k(k\le M)$ 件次品的概率.

**解** 有放回抽样 $n$ 件, 可以看成做 $n$ 次独立试验,E 为取到次品. 每件取到次品的概率为 $p=\dfrac{M}{N}$, 取到正品的概率 $q=1-\dfrac{M}{N}$. 因而, 取到 $k$ 件次品的概率为

$$C_n^k\left(\frac{M}{N}\right)\left(1-\frac{M}{N}\right)^{n-k}.$$

当 $n$ 较大时, 计算可能不便. 例如, 做 $n=100$ 次独立试验,$P(A)=p=0.02$ , 求 $A$ 至少发生 40 次的概率, 应为 $\displaystyle\sum_{k=40}^{100} C_{100}^k 0.02^k 0.98^{100-k}$, 计算复杂. 当 $n$ 很大, $p$ 很小时, 可以利用泊松定理近似计算, 并可查泊松分布表方便计算.

**定理 1.6 (泊松定理)** 设 $n$ 为正整数,$\lambda = np_n$ 为常数, 则对任意正整数 $k$ 有

$$\lim_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

**证明**　$p_n = \dfrac{\lambda}{n}$, 则

$$\begin{aligned}C_n^k p_n^k (1-p_n)^{(n-k)} &= \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\&= \frac{\lambda^k}{k!}\left[1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}.\end{aligned}$$

对任意固定的 $k$, 当 $n\to\infty$, 有

$$1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\to 1, \left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}, \left(1-\frac{\lambda}{n}\right)^{-k}\to 1.$$

因此, $\lim\limits_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \dfrac{\lambda^k}{k!}e^{-\lambda}$.

泊松定理中 $\lambda = np_n$ 为常数, 表示 $n$ 很大时,$p_n$ 一定很小. 实际应用中, 常用泊松近似公式, 当 $n$ 充分大时,$p$ 充分小时, 令 $\lambda = np$, 有

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!}e^{-\lambda}.$$

**例 1.14** 工厂有独立运行的同型设备 80 台, 每台发生故障的概率为 0.01, 发生故障时需一名维修人员排除. 现有两种配备维修人员的方案. (1) 配备 4 人, 每人负责维修 20 台设备; (2) 配备 3 人, 共同负责维修 80 台设备. 分别求设备发生故障而得不到及时维修的概率.

**解**　(1) 设 $X_i$ 为第 $i$ 人负责的 20 台设备中同时发生的故障数, 记 $A_i = X_i \geq 2$, 当 $A_i$ 至少发生一个时, 设备得不到及时维修,$P(X_i = k) = C_{20}^k 0.01^k \cdot 0.99^{20-k}$.

$$P(A_i = P(X_i \geq 2) = 1 - P(X_i \leq 1) = 1 - 0.99^{20} - C_{20}^1 0.01\cdot 0.99^{19} = 0.01686.$$

所求概率

$$P(A_1\cup\cdots\cup A_4) = 1 - P(\bar{A}_1)\cdots P(\bar{A}_4) = 1-(1-0.01686)^4 = 0.06575.$$

(2) 设 $X$ 为 80 台设备中同时发生的故障数,$X\geq 4$ 发生时, 设备得不到及时维修.

$$P(X=k) = C_{80}^k 0.01^k 0.99^{80-k}.$$

所求概率

$$P(X\geq 4) = 1 - P(X\leq 3) = 1 - \sum_{k=0}^{3} C_{80}^k \cdot 0.01^k \cdot 0.99^{80-k} = 0.00866.$$

第二种情况中, 配备人员减少了, 但设备得不到及时维修的概率却下降了. 说明我们可以利用概率论的方法解决实际问题.

# 2　随机变量及其概率分布

## 2.1　随机变量及其分布函数

### 2.1.1　随机变量

随机试验按其结果可以分为两类: 一类与数值直接相关, 另一类与数值没有直接联系. 在第二类情况中, 我们可以通过让随机试验的每一个结果对应一个实数的方法, 人为地使其与数值对应.

对随机试验的每个结果 $e$ 总能用实数 $x$ 与之对应. 由于试验结果的出现是随机的, 困而 $X$ 的取值也是一个随机变化的量. 我们称之为随机变量.

**定义 2.1** 设 $\Omega = e$ 是随机试验的样本空间, 如果对每个 $e \in \Omega$, 都对应一个单值实函数 $X = X(e)$ , 称 $X$ 为**随机变量**.

我们一般以大写字母 $X, Y, Z$ 等表示随机变量, 以小写字母 $x, y, z$ 等表示实数.

随机变量与随机事件的关系: 随机变量在某范围内取值表示随机事件.

有一个常用的随机变量, 用于描述事件 $A$ 发生的情况, 即定义

$$\begin{cases} X = 1, \text{若事件发生} \\ X = 0, \text{若事件不发生} \end{cases}$$

$X$ 可以看成一次试验中事件 $A$ 发生的次数.

引入随机变量后, 我们就可以利用随机变量来描述随机现象, 对事件及事件概率的研究扩大为对随机变量及其取值规律的研究. 有了随机变量, 我们还可以使用更多的数学工具.

### 2.1.2　随机变量的分布函数

对任意实数 $x$ ,$\{X \leq x\}$ 可以表示随机事件, 其概率 $P(X \leq x)$ 与 $x$ 有关, 对每个 $x$ , 都有唯一的概率值 $P(X \leq x)$ 与之对应, 从而构成函数关系, 记为 $F(x) = P(X \leq x)$. 显然, $F(x)$ 是一个取值范围在 $[0, 1]$ 的普通函数.

**定义 2.2** 设 $X$ 是一个随机变量, $x$ 是任意实数, 称函数

$$F(x) = P(X \leq x), \qquad -\infty < x < \infty.$$

为随机变量 X 的分布函数.

**补充**　利用了概率的连续性.$P(\{X \leq x\}) = P(\lim\limits_{\varepsilon \to 0+} \{X \leq x + \varepsilon\}) = \lim\limits_{\varepsilon \to 0+} P(X \leq x + \varepsilon)$.

如果把随机变量的取值看成数轴上的一个点, 则 $F(x)$ 表示随机变量 $X$ 在区间 $(-\infty, x]$ 取值的概率 (图 4).

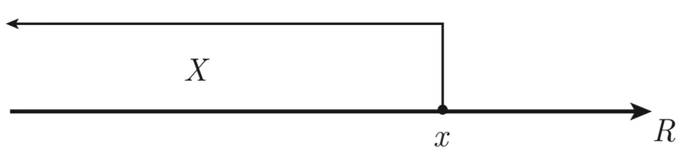


图 4

**例 2.1** 某人射击命中率为 0.6 , 独立射击 2 枪, 命中 $X$ 枪, 求 $X$ 的分布函数.

**解** 独立射击 2 枪, 相当于做 2 次独立重复试验, 命中数 $X$ 的可能取值为 0 , 1 , 2.

$$
\begin{aligned}
&P(X=0)=(1-0.6)^2=0.16,\\
&P(X=1)=C_2^1\cdot 0.6\cdot(1-0.6)=0.48,\\
&P(X=2)=0.6^2=0.36.
\end{aligned}
$$

根据分布函数定义及概率可加性有

当 $x<0$ 时, $X\le x$ 是不可能事件, $F(x)=P(X\le x)=0$;

当 $0\le x<1$ 时,$X\le x\Leftrightarrow X=0, F(x)=P(X\le x)=P(X=0)=0.16$;

当 $1\le x<2$ 时,$X\le x\Leftrightarrow\{X=0\}\cup\{X=1\}$ ,

$$F(x)=P(X\le x)=P(X=0)+P(X=1)=0.16+0.48=0.64.$$

当 $X\ge 2$ 时, $X\le x$ 是必然事件,$F(x)=P(X\le x)=1$.

因此有

$$
F(x)=\begin{cases}0, & x<0\\ 0.16, & 0\le x<1\\ 0.64, & 1\le x<2\\ 1, & x\ge 2\end{cases}
$$

$F(x)$ 图形为阶梯上升函数, 在 0 , 1 , 2 处有跳跃, 见图 5.

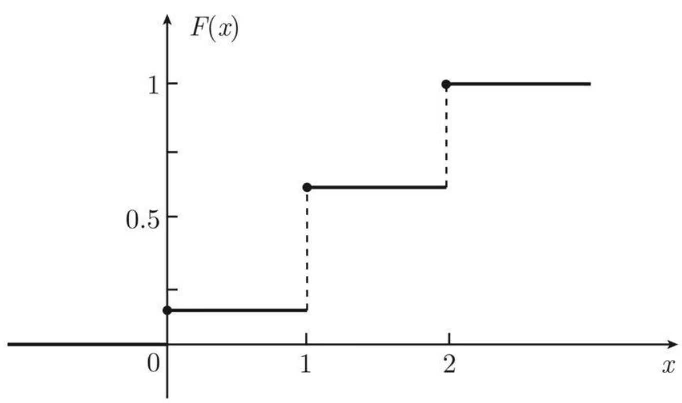


图 5

有了随机变量 $X$ 的分布函数 $F(x)$, 就可以用其描述 $X$ 在某些范围取值的概率, 如对任意实数

$a, b(a < b)$ 有

$$
\begin{aligned}
&P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a), \\
&P(X > b) = 1 - P(X \le b) = 1 - F(b), \\
&P(X \ge b) = 1 - P(X < b) = 1 - F(b\textbf{-0}).
\end{aligned}
$$

**补充** $P(X < x) = P(\lim_{\varepsilon \to 0+} \{X \le x - \varepsilon\}) = \lim_{\varepsilon \to 0+} P(X \le x - \varepsilon) = \lim_{\varepsilon \to 0+} F(x - \varepsilon) = F(x - 0)$.
故

$$
\begin{aligned}
&P(a < X < b) = F(b - 0) - F(a) \\
&P(a \le X < b) = F(b - 0) - F(a - 0) \\
&P(a < X \le b) = F(b - 0) - F(a) \\
&P(a \le X \le b) = F(b) - F(a - 0)
\end{aligned}
$$

分布函数 $F(x)$ 有如下**基本性质**:

(1) $F(x)$ 为单调不减函数, 即对任意实数 $x_2 > x_1$, 有 $F(x_2) > F(x_1)$.

**证明** 对 $x_2 > x_1, F(x_2) - F(x_1) = P(X \le x_2) - P(X \le x_1) = P(x_1 < X \le x_2) \ge 0$.

(2) $0 \le F(x) \le 1$, 且 $F(-\infty) = \lim_{x \to -\infty} F(x) = 0, F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.

**证明** $\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P(X \le x) = P(\lim_{x \to -\infty} \{X \le x\}) = P(\Phi) = 0$.
$\lim_{x \to +\infty} F(x) = \lim_{x \to +\infty} P(X \le x) = P(\lim_{x \to +\infty} \{X \le x\}) = P(\Omega) = 1$.

(3) $F(x)$ 为右连续, 即 $F(x_0 + 0) = \lim_{x \to x_0 + 0} F(x) = F(x_0)$.
左不连续, 即 $F(x) = P(X \le x) \ne \lim_{\varepsilon \to 0+} F(x - \varepsilon) = P(X < x)$.

**证明** $\lim_{x \to x_0 + 0} F(x) = \lim_{x \to x_0 + 0} P(X \le x) = P(\lim_{x \to x_0 + 0} \{X \le x\}) = P(X \le x_0) = F(x_0)$.

任一随机变量的分布函数 $F(x)$ 必满足以上三性质. 反之, 任一函数 $F(x)$ 具有以上三性质, 则 $F(x)$ 必定是某个随机变量的分布函数.

$$
F(x)\text{在}x_0\text{处连续} \Leftrightarrow P(X = x_0) = 0.
$$

**例 2.2** 设随机变量 $X$ 的分布函数为

(1)

$$
F(x) = A + B \arctan x, \qquad -\infty < x < +\infty.
$$

(2)

$$
F(x) = \begin{cases} 0, & x \le -a \\ A + B \arcsin(x/a), & -a < x \le a \\ 1, & x \ge a \end{cases}
$$

试求常数 $A, B$.

**解** (1) 根据分布函数性质 (2),

$$0=\lim_{x\to-\infty}F(x)=\lim_{x\to-\infty}(A+B\arctan x)=A-\frac{\pi}{2}B.$$
$$1=\lim_{x\to+\infty}F(x)=\lim_{x\to+\infty}(A+B\arctan x)=A+\frac{\pi}{2}B.$$

解得

$$A=\frac{1}{2},\quad B=\frac{1}{\pi}.$$

(2) 根据分布函数的右连续性,

$$\begin{cases}\lim\limits_{x\to-a+0}F(x)=F(-a)\\ \lim\limits_{x\to a+0}F(x)=F(a)\end{cases}\qquad \text{即}\begin{cases}A+B\arcsin(-1)=0\\ A+B\arcsin(1)=1\end{cases}$$

得 $A-\frac{\pi}{2}B=0, A+\frac{\pi}{2}B=1$, 解得 $A=\frac{1}{2}, B=\frac{1}{\pi}$.

## 2.2 离散型随机变量及其分布

### 2.2.1 离散型随机变量

本节主要研究一类常用的随机变量, 这类随机变量的可能取值具有特征: 取有限个值 $x_1,x_2,...,x_n$, 或可列无限个值 $x_1,x_2,...,x_n,....$

**定义 2.3** 若随机变量 $X$ 的可能取值为有限个或可列无限多个, 则称 $X$ 为**离散型随机变量**.

非离散型随机变量 $X$ 的所有可能取值不能按一定的顺序排列起来, 不能明确指出某个取值的前或后面是哪一个值. 例如, $X$ 的取值为 $(0,1)$ 区间内的所有实数, 无法指出 0.5 之后 $X$ 取哪个值. 我们研究离散型随机变量, 不仅要知道它取哪些值, 而且要知道这些取值对应的概率. 通过离散型随机变量的分布律可以了解这些信息.

**定义 2.4** 设离散型随机变量 $X$ 的所有可能取值为 $X_1,X_2,...,X_n,...,X$ 取值 $X_k$ 的相应概率为 $P_k$ , 即

$$P(X=x_k)=P_k, k=1,2,...$$

我们称之为离散型随机变量 $X$ 的分布律.

分布律也可以用表格的形式给出:

| $X$ | $x_1$ | $x_2$ | ... | $x_n$ | ... |
|---|---|---|---|---|---|
| $P$ | $p_1$ | $p_2$ | ... | $p_n$ | ... |

表 1

分布律具有如下性质:

(1) $P_k\geq 0, k=1,2,...,$

(2) $\sum\limits_k p_k=1.$

### 2.2.2 常用离散型随机变量

**1. 0 - 1 分布**

若随机变量 $X$ 的可能取值只有 0 , 1 , 且 $P(X=1)=p, P(X=0)=q, 0<p<1, q=1-p$.

或用表格形式,

| $X$ | 0 | 1 |
|---|---|---|
| $P$ | $p$ | $q$ |

表 2

则称 $X$ 服从 0 - 1 分布.

**2. 二项分布**

若随机变量 $X$ 的分布律为

$$P_k=P(X=k)=C_n^k p^k q^{n-k}, k=0,1,...,n,$$

其中 $0<p<1, q=1-p$. 则称 $X$ 服从参数为 $n, p$ 的二项分布, 记为 $X\sim B(n,p)$.

$n=1$ 时, $P(X=k)=p^k q^{1-k}, k=0,1$. 二项分布退化为 0-1 分布.

容易验证二项分布的分布律满足分布律性质, 显然 $P_k\geq 0$. 且

$$\sum_{k=0}^{n} p_k=\sum_{k=0}^{n} C_n^k p^k q^{n-k}=(p+q)^n=1.$$

可见满足分布律性质.

二项分布的随机变量 $X$ 的概率背景: 设 $n$ 重伯努利试验中 $A$ 发生的次数为 $X$ , $A$ 发生的概率为 $p$ , 则 $X$ 服从二项分布 $B(n,p)$.

下面以 $X\sim B(10,0.6)$ 为例, 我们通过图表形式看一下 $X$ 的概率分布情况. 由图 6 可见, 对固定的 $n$ 和 $p$ ,$X=k$ 的概率先随 $k$ 的增加而增加, 达到最大值后, 再单调减小. 是一个**单峰函数**.

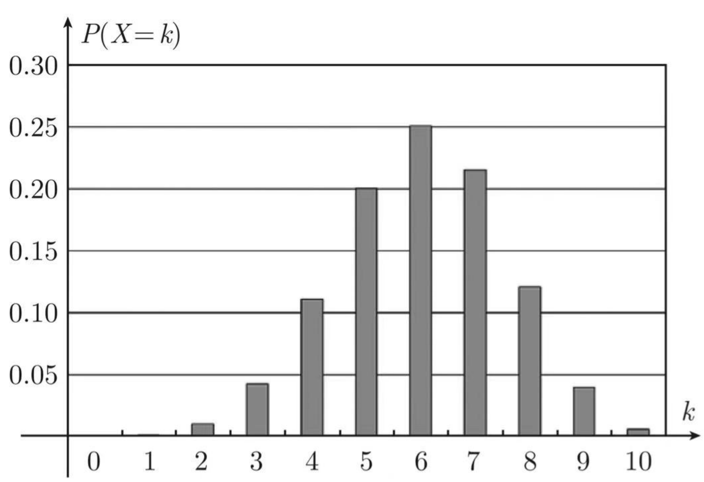


图 6

以下讨论, 对 $X \sim B(n,p)$ , $k$ 为何值时 $P(X=k)$ 达到最大值? 设 $P_k = P(X=k)$ 在 $k=k_0$ 达到最大值, 则

$$\frac{p_{k_0}}{p_{k_0-1}} = \frac{C_n^{k_0} p^{k_0} q^{n-k_0}}{C_n^{k_0-1} p^{k_0-1} q^{n-k_0+1}} = \frac{(n-k_0+1)p}{k_0 q} \geq 1.$$
$$\frac{p_{k_0}}{p_{k_0+1}} = \frac{C_n^{k_0} p^{k_0} q^{n-k_0}}{C_n^{k_0+1} p^{k_0+1} q^{n-k_0-1}} = \frac{(k_0+1)p}{(n-k_0)p} \geq 1.$$

解出 $(n+1)p - 1 \leq k_0 \leq (n+1)p$.

由此得到结论: (1) 当 $(n+1)p$ 为整数时,$P_k$ 在 $k=(n+1)p-1$ 和 $k=(n+1)p$ 达到最大.

(2) 当 $(n+1)p$ 不是整数时,$P_k$ 在 $k=[(n+1)p]$ 达到最大. 其中, $[x]$ 表示不超过 $x$ 的最大整数.

例如,$X \sim B(14, 0.4)$ , $(n+1)p = 6$, 所以 $p_k$ 在 $k=5$ 和 $k=6$ 处达到最大值 0.2066.

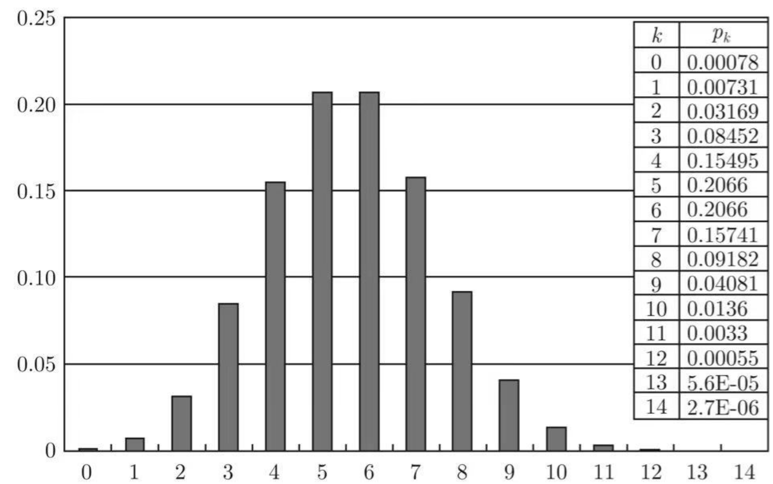


图 7

**3. 泊松分布**

如果随机变量 $X$ 的分布律为

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

其中 $\lambda > 0$ 为常数, 则称 $X$ 服从参数为 $\lambda$ 的泊松分布, 记为 $X \sim P(\lambda)$.

验证分布律性质,

显然 $P_k \geq 0$, 且 $\sum\limits_{k=0}^{\infty} p_k = \sum\limits_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum\limits_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$.

满足分布律性质.

泊松分布是概率论中的重要分布之一, 常用于描述大量试验中稀有事件出现频数的概率模型. 例如, 单位时间内电话总机接到的呼叫数, 每天某高速公路发生的交通事故数, 每页印刷品出现的印刷错误数, 这些都服从泊松分布. 还有很多其他的实例.

**例 2.3** 设随机变量 $X$ 服从参数为 $\lambda$ 的泊松分布, 且已知 $P(X=1) = P(X=2)$ , 求 $P(X=4)$.

**解** $X$ 的分布律为

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

由条件 $P(X=1)=P(X=2)$ , 得到

$$\frac{\lambda^1}{1!}e^{-\lambda}=\frac{\lambda^2}{2!}e^{-\lambda}.$$

解得 $\lambda=2$ , 因此

$$P(X=4)=\frac{2^4}{4!}e^{-2}=\frac{2}{3}e^{-2}.$$

**4. 几何分布**

如果随机变量 $X$ 的分布律为

$$p_k=P(X=k)=q^{k-1}p,\quad k=1,2,...$$

其中 $0<p<1$ 为常数, $q=1-p$ , 则称 $X$ 服从参数为 $p$ 的几何分布, 记为 $X\sim g(P)$.

验证分布律性质, 显然 $P_k\geq 0$, 且

$$\sum_{k=1}^{\infty}p_k=\sum_{k=1}^{\infty}q^{k-1}p=p\sum_{k=1}^{\infty}q^{k-1}=p\frac{1}{1-q}=1.$$

满足分布律性质.

几何分布的概率背景: 在 $n$ 重伯努利试验中, 每次只有两个结果 $A$ 与 $\bar{A}$, 设 $P(A)=p$ . 将试验进行到 $A$ 首次出现为止. 令 $X$ 为 $A$ 首次出现时所需的试验次数, 则 $X$ 服从几何分布 $g(p)$.

**几何分布的无记忆性** 在 $n$ 重伯努利试验中, $A$ 首次出现时所需的试验次数服从几何分布, 若前 $t$ 次试验 $A$ 没有出现, 则再试验 $s$ 次. $A$ 首次出现的概率与前面没有成功的 $t$ 次试验无关. 好像把前面 $t$ 次试验给忘了, 这称为无记忆性. 无记忆性的概率表达为 $P(X=s+t\mid X>t)=P(X=s)$.

**证明** $X\sim g(p)$, 则 $P(X=k)=q^{k-1}p$ , 其中 $q=1-p,k=1,2,...$

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

$$\begin{aligned}P(X=s+t\mid X>t)&=\frac{P(X=s+t\cap X>t)}{P(X>t)}=\frac{P(X=s+t)}{P(X>t)}\\&=\frac{q^{s+t-1}p}{q^t}=q^{s-1}p=P(X=s).\end{aligned}$$

得证.

## 2.3 连续型随机变量及其分布

### 2.3.1 连续型随机变量

有一类随机变量, 其取值可以取到某区间内的所有实数, 不能一一列举. 如元件的寿命, 等候班车的时间, 工件的长度等. 以下我们给出连续型随机变量的定义. **定义 2.5** 设随机变量 $X$ 的分布函数为 $F(x)$ , 若存在非负可积函数 $p(x)$ , 对任意实数 $x$ 有

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

则称 $X$ 为**连续型随机变量**, 称函数 $p(x)$ 为 $X$ 的概率密度函数, 简称密度函数.

由定义可见, 任何连续型随机变量的密度函数具有如下性质:

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

(2) $\displaystyle\int_{-\infty}^{+\infty} p(x)\mathrm{d}x=1$.

(3) 对于任意实数 $a,b(a<b)$,

$$P(a<X\leq b)=F(b)-F(a)=\int_{a}^{b} p(x)\mathrm{d}x.$$

(4) 由于 $p(x)$ 为可积函数, 根据微积分性质, 分布函数 $F(x)$ 为连续函数.

(5) 若 $p(x)$ 在点 $x$ 连续, 则分布函数 $F(x)$ 在 $x$ 处可导, 且 $p(x)=F'(x)$.

关于 (2) 的说明, $\displaystyle\int_{-\infty}^{+\infty} p(x)\mathrm{d}x=\lim_{x\to+\infty}\int_{-\infty}^{x} p(x)\mathrm{d}x=\lim_{x\to+\infty} F(x)=1$.

另外, 从图形上看, (2) 表示函数 $p(x)$ 在整个 $x$ 轴上覆盖的面积为 1(图 8). (3) 表示随机变量 $X$ 在区间 $(a,b]$ 取值的概率等于密度函数 $p(x)$ 在 $(a,b)$ 覆盖的曲边梯形面积. 后面会证明, 这一结论与区间的开闭无关.

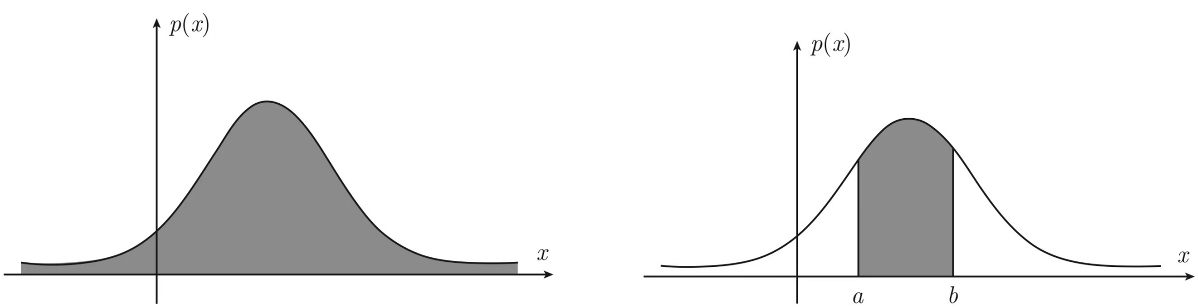


图 8

在性质 (3) 中, 令 $a=x,b=x+\Delta x$ , 则有

$$P(x<X\leq x+\Delta x)=F(x+\Delta x)-F(x)=\int_{x}^{x+\Delta x} p(x)\mathrm{d}x\approx p(x)\Delta x.$$

这说明, 连续型随机变量 $X$ 在小区间 $\Delta x$ 内取值的概率近似与密度成正比, 一点 $x$ 处的密度函数值越大, 该点附近取值的概率也越大. 这一点与物理学中的密度性质相似, 这也就是 $p(x)$ 称为概率密度

函数的原因.

连续型随机变量 $X$ 有一重要特性, 即 $X$ 在一点 $x_0$ 处取值的概率为 0. 事实上, 对任意 $\Delta x > 0$ 有

$$0 \le P(X = x_0) \le P(x_0 - \Delta x < X \le x_0) = F(x_0) - F(x_0 - \Delta x).$$

由于分布函数 $F(x)$ 为连续函数, 令 $\Delta x \to 0$ , 上式趋向于 0 , 所以 $P(X = X_0) = 0.$

根据这一性质我们可知, 连续型随机变量 $X$ 在某一区间取值的概率与区间的开或闭无关, 所以

$$P(a < X < b) = P(a < X \le b) = P(a \le X < b) = P(a \le X \le b) = \int_a^b p(x)\mathrm{d}x.$$

对连续型随机变量 $X$ , 在一点取值的概率为 0 , 即 $P(X = X_0) = 0$ . 但是事件 $\{X = X_0\}$ 是有可能发生的. 因此对连续型随机变量, 概率为 0 的事件与不可能事件并不等价.

**例 2.4** 设随机变量 $X$ 的概率密度函数为

$$p(x) = Ae^{-|x|}, \quad -\infty < x < \infty.$$

试求, (1) 常数 $A$; (2)$P(0 < X < 2)$; (3) 分布函数 $F(x)$.

**解** (1) $\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = \int_{-\infty}^{0} Ae^{x}\mathrm{d}x + \int_{0}^{+\infty} Ae^{-x}\mathrm{d}x = 2A = 1$. 因此 $A = \frac{1}{2}$. 从而,$p(x) = \frac{1}{2}e^{-|x|}, -\infty < x < \infty$.

(2) $P(0 < X < 2) = \int_0^2 \frac{1}{2}e^{-x}\mathrm{d}x = \frac{1}{2}(1 - e^{-2})$.

(3)$x < 0$ 时, $F(x) = \int_{-\infty}^{x} p(x)\mathrm{d}x = \int_{-\infty}^{x} \frac{1}{2}e^{x}\mathrm{d}x = \frac{1}{2}e^{x}$.

$x \ge 0$ 时, $F(x) = \int_{-\infty}^{x} p(x)\mathrm{d}x = \int_{-\infty}^{0} \frac{1}{2}e^{x}\mathrm{d}x + \int_{0}^{x} \frac{1}{2}e^{-x}\mathrm{d}x = 1 - \frac{1}{2}e^{-x}$.

因此

$$F(x) = \begin{cases} \frac{1}{2}, & x < 0 \\ 1 - \frac{1}{2}e^{-x}, & x \ge 0 \end{cases}$$

利用 $F(x)$, 也可求 $P(0 < X < 2) = F(2) - F(0) = \left(1 - \frac{1}{2}e^{-2}\right) - \frac{1}{2} = \frac{1}{2}(1 - e^{-2})$.

### 2.3.2 常用连续型随机变量

**1. 均匀分布**

若随机变量 $X$ 的概率密度函数为

$$p(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, \text{其他} \end{cases}$$

则称 $X$ 在区间 $[a, b]$ 上服从均匀分布, 记为 $X \sim U[a, b]$.

$X$ 的分布函数为

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

均匀分布的密度函数 $p(x)$ 及分布函数 $F(x)$ 图形如图 9 所示.

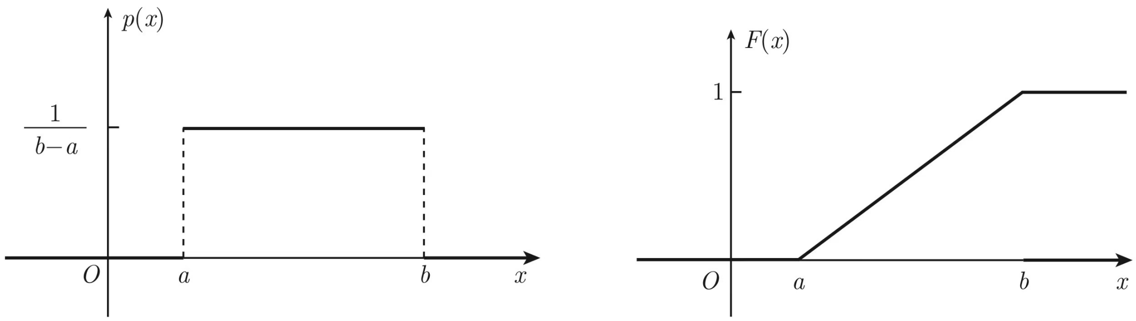


图 9

设随机变量 $X$ 服从 $[a,b]$ 上的均匀分布, 考虑 $X$ 在 $[a,b]$ 内的子区间 $(x,x+\Delta x)$ 取值的概率

$$P(x<X<x+\Delta x)=\int_{x}^{x+\Delta x} p(x) \mathrm{d} x=\int_{x}^{x+\Delta x} \frac{1}{b-a} \mathrm{~d} x=\frac{\Delta x}{b-a}.$$

此概率为子区间与原区间的长度之比, 与子区间的位置无关. 困此只要子区间长度相同,$X$ 落在其中的概率是等可能的. 这也体现了均匀分布的**均匀性**.

**2. 指数分布**

若随机变量 $X$ 的概率密度函数为

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, x \geq 0 \\ 0, x<0\end{cases}$$

其中 $\lambda>0$ 为常数, 则称 $X$ 服从参数为 $\lambda$ 的指数分布, 记为 $X \sim E(\lambda)$.

$X$ 的分布函数为

$$p(x)=\begin{cases}1-e^{-\lambda x}, x \geq 0 \\ 0, \text{其他}\end{cases}$$

指数分布的密度函数 $p(x)$ 及分布函数 $F(x)$ 的图形如图 10 所示.

指数分布有很多实际应用, 常用于描述各种寿命的分布, 如电子设备的寿命, 生物体的寿命, 随机服务系统的服务时间.

**例 2.5** 设某类电子元件的使用寿命 $X$(小时) 服从参数为 $\lambda=1/2000$ 的指数分布,

(1) 任取一只元件, 求能使用 1000 小时以上的概率; (2) 一只元件已经正常使用了 1000 小时以上, 求还能再使用 1000 小时以上的概率.

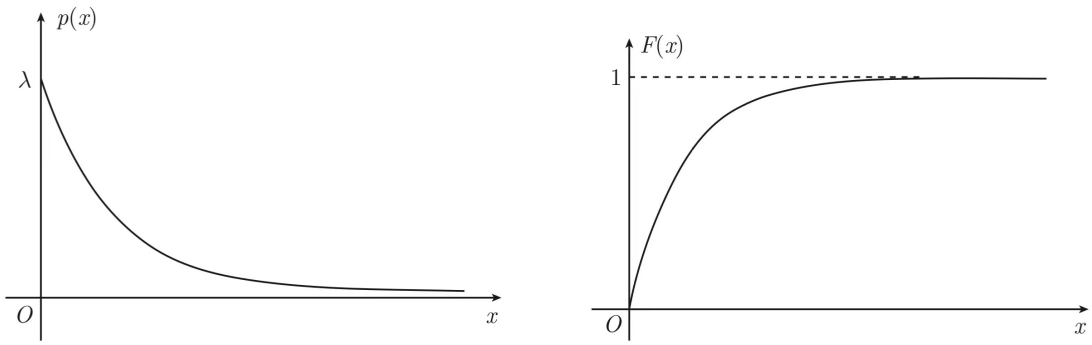


图 10

**解** $X$ 的分布函数为

$$p(x) = \begin{cases} 1 - e^{-\frac{1}{2000}x}, x \geq 0 \\ 0, x < 0 \end{cases}$$

(1) $P(X > 1000) = 1 - P(X < 1000) = 1 - F(1000) = 1 - (1 - e^{-1/2}) = e^{-1/2}$.

(2) 所求概率为

$$\begin{aligned} P(X > 2000 \mid X > 1000) &= \frac{P(X > 2000 \cap X > 1000)}{P(X > 1000)} = \frac{P(X > 2000)}{P(X > 1000)} \\ &= \frac{1 - F(2000)}{1 - F(1000)} = \frac{e^{-1}}{e^{-1/2}} = e^{-1/2}. \end{aligned}$$

一只元件能使用 1000 小时以上的概率为 $e^{-1/2}$, 已使用了 1000 小时以上, 再使用 1000 小时以上的概率仍然为 $e^{-1/2}$ , 这反映了指数分布的一个重要特性: **无记忆性**. 即已使用了 1000 小时以上, 再使用 1000 小时以上的概率不受影响, 好像忘记了已使用 1000 小时以上.

无记忆性的一般表示为

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

实际上,

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(\{X > s + t\} \cap \{X > s\})}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{1 - F(s + t)}{1 - F(s)} = \frac{e^{-\lambda(s+t)}}{e^{\lambda s}} = e^{-\lambda t} = P(X > t). \end{aligned}$$

**3. 正态分布**

若随机变量 $X$ 的密度函数为

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

其中 $\mu, \sigma(> 0)$ 为常数, 则称 $X$ 服从参数为 $\mu, \sigma^2$ 的正态分布, 记为 $X \sim N(\mu, \sigma^2)$.

正态分布的分布函数为

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

正态分布的密度函数 $p(x)$ 及分布函数 $F(x)$ 的图形如图 11 所示.

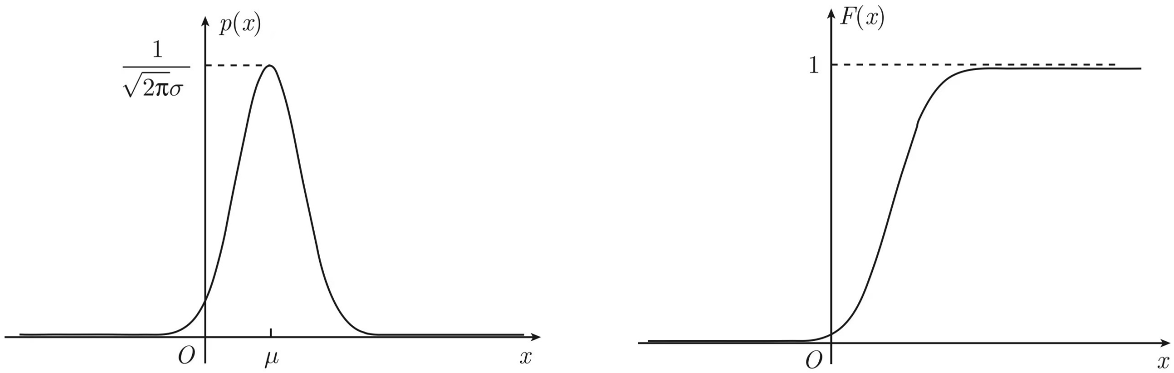


图 11

利用极坐标和二重积分, 可以验证 $\displaystyle\int_{-\infty}^{\infty} p(x)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t = 1.$

**证明** 先化归为标准正态分布 $X \sim N(0,1)$.

$$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x \quad \text{是不可积的式子, 不妨转化为二重积分.}$$

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \mathrm{d}y &= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \mathrm{d}x\mathrm{d}y \\
&\xlongequal[y=r\sin\theta]{x=r\cos\theta} \int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi} e^{-\frac{r^2}{2}} r\mathrm{d}r\mathrm{d}\theta \\
&= \int_0^{\infty} -e^{-\frac{r^2}{2}} \mathrm{d}\left(\frac{r^2}{2}\right) = 1.
\end{aligned}$$

正态分布 $X \sim N(\mu, \sigma^2)$ 密度函数 $p(x)$ 的图形具有以下特点:

(1) 曲线 $p(x)$ 关于 $x = \mu$ 对称, 且对任意 $b > 0$, 有 $P(X \leq \mu - b) = P(X \geq \mu + b)$.

(2) $x \to \pm\infty, p(x) \to 0$ , 曲线以 $x$ 轴为渐近线.

(3) $x = \mu$ 时 $p(x)$ 取到最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$, 由于密度图形覆盖面积总为 1 , 因此固定 $\mu$ 时, $\sigma$ 越小, 最大值越大, 图形越高越陡峭;$\sigma$ 越大, 最大值越小, 图形越低越平缓.

(4) 固定 $\sigma, \mu$ 变化时, 曲线沿对称轴 $x = \mu$ 平移.

当 $\mu = 0, \sigma = 1$ 时,$X \sim N(0,1)$ , 称 $X$ 服从标准正态分布, 其密度函数和分布函数特别记为 $\phi(x)$ 和 $\Phi(x)$.

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

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

事实上, 性质 (1) 中令 $\mu = 0$, 有 $P(X \leq -b) = P(X \geq b)$, 从而

$$\Phi(-x) = P(X \leq -x) = P(X \geq x) = 1 - P(X < x) = 1 - \Phi(x).$$

正态分布是概率统计中最重要的分布之一, 很多自然和社会现象中的随机变量都服从正态分布, 如某地区的成年男子的身高, 一批工件的直径, 测量误差等. 正态分布的概率计算常通过查表实现, 即将一般的正态分布化为标准正态分布, 然后查标准正态分布的分布函数表 $\Phi(x)$ , 得到所求概率.

**定理 2.1** 若随机变量 $X \sim N(\mu, \sigma^2)$, 则 $Z = \dfrac{X-\mu}{\sigma} \sim N(0, 1)$.

**证明** 只要证明 $Z$ 的分布函数为 $\Phi(x)$ 即可. $Z$ 的分布函数为

$$P(Z \leq X) = P(\frac{X-\mu}{\sigma} \leq x) = P(X \leq \mu + \sigma x) = \int_{-\infty}^{\mu+\sigma x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t.$$

令 $\dfrac{t-\mu}{\sigma} = v$, 得到,$\displaystyle\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \mathrm{d}v$, 得证.

由定理, 若 $X \sim N(\mu, \sigma)$, 则 $X$ 的分布函数 $F(x)$ 和概率 $P(a < X < b)$ 可用 $\Phi(x)$ 表示:

$$F(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$
$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

利用上式, 我们可以得到,

$$P(\mid X-\mu \mid \leq \sigma) = P\left(-1 \leq \frac{X-\mu}{\sigma} \leq 1\right) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826$$
$$P(\mid X-\mu \mid \leq 2\sigma) = P\left(-2 \leq \frac{X-\mu}{\sigma} \leq 2\right) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0.9544$$
$$P(\mid X-\mu \mid \leq 3\sigma) = P\left(-3 \leq \frac{X-\mu}{\sigma} \leq 3\right) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 0.9974.$$

我们看到, 正态分布的随机变量 $X \sim N(0, 1)$ , 虽然其理论取值范围是 $(-\infty, +\infty)$, 但是在区间 $[\mu - 3\sigma, \mu + 3\sigma]$ 内的取值的可能性超过 $99.7\%$ , 在这之外的取值的可能性不足千分之三. 这表明在一次随机试验中, X 的取值几乎可以肯定落入 $[\mu - 3\sigma, \mu + 3\sigma]$ 中, 这称为"$3\sigma$ 法则'. 该法则在针对实际问题, 特别是产品质量问题的统计推断中有着重要的应用.

**例 2.6** 某单位计划招聘 155 名员工, 按考试成绩从高分到低分依次录取. 现有 526 人报名, 假设报名者的成绩 $X \sim N(\mu, \sigma^2)$. 已知报名者中 90 分以上有 12 人, 60 分以下有 83 人. 已知某人成绩为 78

分, 问此人能否被录取?

**解** $X \sim N(\mu, \sigma^2), \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

90 分以上有 12 人, 因此

$$P(X > 90) = \frac{12}{526} = 0.02281, P(X \le 90) = 1 - 0.0228 = 0.9772.$$

$$P(X \le 90) = P\left(\frac{X-\mu}{\sigma} \le \frac{90-\mu}{\sigma}\right) = \Phi(\frac{90-\mu}{\sigma}) = 0.9772, \text{查表}, \frac{90-\mu}{\sigma} = 2.$$

60 分以下有 83 人, 因此

$$P(X < 60) = \frac{83}{526} = 0.1578.$$

$$P(X < 60) = P\left(\frac{X-\mu}{\sigma} < \frac{60-\mu}{\sigma}\right) = \Phi(\frac{60-\mu}{\sigma}) = 0.1578, \text{查表}, -\frac{60-\mu}{\sigma} = 1.$$

解 $\mu, \sigma$ 的方程组, 得到 $\mu = 70, \sigma = 10$. 因此, $X \sim N(70, 10^2)$.

$$\begin{aligned} P(X \ge 78) &= 1 - P(X < 78) = 1 - P\left(\frac{X-70}{10} < \frac{78-70}{10}\right) = 1 - \Phi(0.8) \\ &= 1 - 0.7881 = 0.2119. \end{aligned}$$

而录取率为 $\dfrac{155}{526} = 0.2947 > 0.2119$ , 所以此人能被录取.

## 2.4 随机变量函数的分布

设 $X$ 是随机变量, $y = g(x)$ 是普通实函数. 构造随机变量 $Y$ , 当 $X$ 取值 $x$ 时, $Y$ 取值 $y = g(x)$ , 称 $Y$ 是随机变量 $X$ 的函数, 记为 $Y = g(X)$. 我们需要讨论, 已知 $X$ 的概率分布, 如何求 $Y = g(X)$ 的概率分布. 分 $X$ 为离散型和连续型两种情况讨论.

### 2.4.1 离散型随机变量函数

设 $X$ 是离散型随机变量, 其分布律为

| $X$ | $x_1$ | $x_2$ | ... | $x_n$ | ... |
|---|---|---|---|---|---|
| $P$ | $p_1$ | $p_2$ | ... | $p_n$ | ... |

表 3

对于 $X$ 的函数 $Y = g(X)$ , $Y$ 也是离散型随机变量, 求 $Y$ 的分布律.

当 $X$ 取值 $x_k$ 时, $Y = g(X)$ 取值 $y_k = g(x_k), k = 1, 2, ...$

(1) 若 $y_1, y_2, ..., y_n, ...$ 的取值互不相同, 则 $X = X_k$ 时, 一定有 $Y = y_k = g(x_k)$. 所以 $Y = y_k$ 的概率与 $X = x_k$ 的概率相同, 即

$$P(Y = y_k) = P(Y = g(x_k)) = P(X = x_k) = p_k.$$

从而得到 $Y$ 的分布律:

| $y$ | $y_1$ | $y_2$ | ... | $y_n$ | ... |
|---|---|---|---|---|---|
| $P$ | $p_1$ | $p_2$ | ... | $p_n$ | ... |

表 4

(2) 若 $y_1, y_2, ..., y_n, ...$ 的取值有相同的, 则应把相同的 $y_k$ 对应的概率相加, 作为 $P(Y = y_k)$ 的概率取值. 即

$$P(Y = y_k) = \sum_{i: y_k = g(x_i)} P(X = x_i).$$

同样得到 $Y$ 的分布律.

### 2.4.2 连续型随机变量函数

设 $X$ 是连续型随机变量,$y = g(x)$ 为连续实函数, 则 $Y = g(X)$ 也是连续型随机变量. 已知 $X$ 的密度函数 $p_X(x)$ , 求 $Y$ 的密度函数 $p_Y(y)$.

常用的方法是分布函数法, 即先求 $Y$ 的分布函数 $F_Y(y)$.

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{x: g(x) \leq y} p_X(x) \mathrm{d}x.$$

然后, $p_Y(y) = F'_Y(y)$.

**例 2.7** 设随机变量 $X$ 服从 $[-1, 1]$ 上的均匀分布,$Y = 2X + 1$, 求 $Y$ 的密度函数 $p_Y(y)$.

**解** $X$ 的密度函数

$$p(x) = \begin{cases} 1/2, -1 \leq x \leq 1 \\ 0, \text{其他} \end{cases}$$

记 $Y$ 的分布函数为 $F_Y(y)$ , 由于 $Y$ 的可能取值范围是 $[-1, 3]$, 因此

$y < -1$ 时,

$$F_Y(y) = P(Y \leq y) = 0.$$

$-1 \leq y < 3$ 时,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(2X + 1 \leq y) \\ &= P\left(X \leq \frac{1}{2}y - \frac{1}{2}\right) = \int_{-1}^{\frac{1}{2}y - \frac{1}{2}} \frac{1}{2} \mathrm{d}x = \frac{1}{4}(y + 1). \end{aligned}$$

$y > 3$ 时,

$$F_Y(y) = P(Y \leq y) = 1.$$

得到

$$p_Y(y) = F'_Y(y) \begin{cases} \frac{1}{4}, -1 \leq y \leq 3 \\ 0, \text{其他}. \end{cases}$$

$Y$ 仍然服从均匀分布, $Y \sim U[-1,3]$.

**例 2.8** 设随机变量 $X$ 的概率密度为 $p_X(x)$ ,$-\infty < x < \infty$, 求 $Y = X^2$ 的密度函数 $p_Y(y)$.

**解** 先求 $Y$ 的分布函数 $F_Y(y)$. $Y$ 的可能取值范围 $[0,+\infty)$, 因此

$y < 0$ 时,

$$F_Y(y) = P(Y \leq y) = 0.$$

$y > 0$ 时,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

对 $F_Y(y)$ 求导,

$$p_Y(y) = F_Y'(y) \begin{cases} \dfrac{1}{2\sqrt{y}}[p_X(\sqrt{y}) + p_X(-\sqrt{y})], y > 0 \\ 0, y \leq 0. \end{cases}$$

若已知 $X$ 的概率密度代入上式, 便可求 $Y = X^2$ 的概率密度, 例如, X 服从标准正态分布, $X$ 的概率密度 $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 代入可得

$$p_Y(y) \begin{cases} \dfrac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, y > 0 \\ 0, y \leq 0. \end{cases}$$

称 $Y$ 服从自由度为 1 的 $\chi^2$ 分布, $\chi^2$ 分布是统计学中的一个重要分布.

在 $y = g(x)$ 严格单调的情况下, 可以直接推导出 $Y = g(X)$ 的密度函数 $p_Y(y)$.

**定理 2.2** 设随机变量 $X$ 的可能取值范围为 $(a,b)$ , $X$ 的概率密度为 $p_X(x)$ ,$a < x < b$(其中 $a$ 可为 $-\infty$ , $b$ 可为 $+\infty$), 设函数 $y = g(x)$ 处处可导, 且恒有 $g'(x) > 0$[或恒有 $g'(x) < 0$] , 则 $Y = g(X)$ 为连续型随机变量, 其概率密度为

$$p_Y(y) \begin{cases} p_X[g^{-1}(y)] \cdot \mid [g^{-1}(y)]' \mid, \alpha < y < \beta \\ 0, \text{其他}. \end{cases}$$

其中, $\alpha = \min\{g(a), g(b)\}, \beta = \max\{g(a), g(b)\}$, $g^{-1}(y)$ 为 $y = g(x)$ 的反函数.

**证明** 设随机变量 $X, Y$ 的分布函数分别为 $F_X(x), F_Y(y)$, 由题设, $X$ 在 $(a,b)$ 取值, 相应 $Y$ 在 $(\alpha,\beta)$ 取值.

(1) 设 $g'(x) > 0, y = g(x)$ 是严格单调递增函数, 其反函数 $x = g^{-1}(y)$ 也在 $(\alpha,\beta)$ 严格单调递增, 且 $[g^{-1}(y)]' > 0$.

当 $y \in (\alpha,\beta)$ 时,

$$g(X) \leq y \Leftrightarrow X \leq g^{-1}(y).$$

因此,

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

所以 $Y$ 的密度

$$\begin{aligned} p_Y(y) = F'_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}(F_X[g^{-1}(y)]) = F'_X(g^{-1}(y)) \cdot [g^{-1}(y)]' \\ &= p_X(g^{-1}(y)) \cdot (g^{-1}(y))' = p_X(g^{-1}(y)) \cdot \mid (g^{-1}(y))' \mid, \alpha < y < \beta. \end{aligned}$$

(2) 设 $g'(x) < 0, y = g(x)$ 是严格单调递减函数, 其反函数 $x = g^{-1}(y)$ 也在 $(\alpha, \beta)$ 严格单调递减, 且 $[g^{-1}(y)]' < 0$.

当 $y \in (\alpha, \beta)$ 时,

$$g(X) \leq y \Leftrightarrow X \geq g^{-1}(y).$$

因此,

$$F_Y(y) = P(Y \leq y) = P(g(X) \geq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

得到 $Y$ 的密度

$$\begin{aligned} p_Y(y) = F'_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}(1 - F_X[g^{-1}(y)]) = -F'_X(g^{-1}(y)) \cdot [g^{-1}(y)]' \\ &= -p_X(g^{-1}(y)) \cdot (g^{-1}(y))' = p_X(g^{-1}(y)) \cdot \mid (g^{-1}(y))' \mid, \alpha < y < \beta. \end{aligned}$$

定理得证.

**例 2.9** 设随机变量 $X \sim N(\mu, \sigma^2)$ , 求 $X$ 的线性函数 $Y = aX + b(a \neq 0)$ 的密度函数 $p_Y(y)$.

**解** $X$ 的密度函数为

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty.$$

由 $Y = aX + b$ 严格单调, 解得反函数

$$x = g^{-1}(y) = \frac{y-b}{a}, [g^{-1}(y)]' = \frac{1}{a}$$

根据定理 2.2 , $Y = aX + b$ 的密度为

$$\begin{aligned} p_Y(y) &= p_X(g^{-1}(y)) \cdot \left|[g^{-1}(y)]'\right| = p_X\left(\frac{y-b}{a}\right) \cdot \left|\frac{1}{a}\right| \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} \cdot \left|\frac{1}{a}\right| = \frac{1}{\sqrt{2\pi}\,|a|\,\sigma} e^{-\frac{[y-(a\mu+b)]^2}{2a^2\sigma^2}}. \end{aligned}$$

可见, $Y$ 仍然服从正态分布, $Y \sim N(a\mu + b, a^2\sigma^2)$.

# 3 二维随机变量及其分布

实际问题中, 很多随机试验的结果是多因素的, 需要两个或多个随机变量来描述, 例如, 在平面上随机投点, 落点坐标需要 $X, Y$ 两个随机变量来确定; 任取一名高考考生, 其语文, 数学, 外语三门课的成绩需要 $X, Y, Z$ 三个随机变量来确定. 研究这些随机变量, 不但要研究单个的 $X, Y, Z$, 还要研究它们之间的相互关系. 这样, 便形成了作为整体考虑的二维和三维随机变量 $(X, Y)$ 和 $(X, Y, Z)$, 多维随机变量也称为随机向量. 由二维随机变量到多维随机变量, 其性质没有本质的不同, 因此本章我们主要研究二维随机变量.

## 3.1 二维随机变量的分布函数

**定义 3.1** 设 $(X, Y)$ 为二维随机变量, 对任意实数 $(x, y)$, 二元函数

$$F(x, y) = P(X \leq x) \cap (Y \leq y) = PX \leq x, Y \leq y.$$

称为二维随机变量 $(X, Y)$ 的联合分布函数, 简称分布函数.

如果将 $(X, Y)$ 看成平面上随机点的坐标, 则分布函数 $F(x, y)$ 在 $(x, y)$ 处的函数值

$$F(x, y) = PX \leq x, Y \leq y.$$

就是随机点 $(X, Y)$ 落入以 $(x, y)$ 为右上顶点的无穷矩形区域的概率 (图 12).

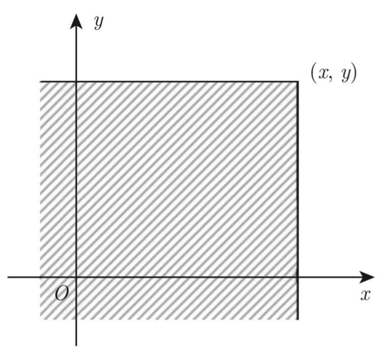


图 12

联合分布函数 $F(x, y)$ 具有以下基本性质:

(1) $F(x, y)$ 分别是 $x$ 和 $y$ 的单调不减函数, 即对任意固定的 $y$, 当 $x_2 > x_1$ 时,$F(x_2, y) \geq F(x_1, y)$; 对任意固定的 $x$, 当 $y_2 > y_1$ 时, 有 $F(x, y_2) \geq F(x, y_1)$.

(2)$0 \leq F(x, y) \leq 1$, 且对任意固定的 $y$ ,$F(-\infty, y) = 0$; 对任意固定的 $x$, $F(x, -\infty) = 0$, 以及 $F(-\infty, -\infty) = 0, F(+\infty, +\infty) = 1$.

(3)$F(x, y)$ 关于 $x, y$ 右连续, 即

$$F(x+0, y) = F(x, y), \quad F(x, y+0) = F(x, y).$$

(4) 对任意实数 $x_2 > x_1, y_2 > y_1$, 有

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0.$$

性质 (1)-性质 (3) 与一维随机变量相似, 性质 (4) 反映了随机变量 $(X, Y)$ 在矩形区域取值的概率非负, 实际上

$$\begin{aligned}
&P(x_1 < X \leq x_2, y_1 < Y \leq y_2)\\
=&P(X \leq x_2, Y \leq y_2) - P(X \leq x_1, Y \leq y_2) - P(X \leq x_2, Y \leq y_1) + P(X \leq x_1, Y \leq y_1)\\
=&F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0.
\end{aligned}$$

任意二维随机变量 $(X, Y)$ 具有性质 (1)-性质 (4), 反之, 如果存在二维函数 $F(x, y)$ 满足以上四条性质, 则必存在随机变量 $X, Y$, 使得 $F(x, y)$ 成为 $(X, Y)$ 的分布函数.

对于二维随机变量 $(X, Y)$, 其中, 单个随机变量 $X$ 或 $Y$ 的分布称为边缘分布. 若已知二维随机变量 $(X, Y)$ 的联合分布函数 $F(x, y)$, 可以导出 $X$ 和 $Y$ 的边缘分布函数 $F_X(x), F_Y(y)$:

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

二维分布函数的概念可以推广到 $n$ 维随机变量.

**定义 3.2** 设 $X_1, X_2, ..., X_n$ 是 $n$ 维随机变量, 对任意实数 $x_1, x_2, ..., x_n$, 称 $n$ 维函数

$$F(x_1, x_2, ..., x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n\}.$$

为随机变量 $(X_1, X_2, ..., X_n)$ 的联合分布函数.

与二维情况类似, 我们可以求得 $(X_1, X_2, ..., X_n)$ 的 $k$ 维边缘分布函数 $(1 \leq k < n)$, 例如,$(X_1, X_2, ..., X_n)$ 中 $X_1$ 的边缘分布函数为

$$F_{X_1}(x_1) = F(x_1, +\infty, ..., +\infty),$$

$(X_1, X_2, ..., X_n)$ 中 $(X_1, X_2)$ 的边缘分布函数为

$$F_{(X_1, X_2)}(x_1, x_2) = F(x_1, x_2, +\infty, ..., +\infty).$$

下面我们通过随机事件的独立性概念引入随机变量相互独立的概念.

**定义 3.3** 设随机变量 $X, Y$ 满足, 对任意 $x, y$, 随机事件 $X \leq x$ 与 $Y \leq y$ 独立, 即

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

即

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

则称随机变量 $X,Y$ 相互独立.

我们不加证明地给出随机变量独立性的一个重要性质: 若随机变量 $X,Y$ 相互独立, $f(x)$,$g(y)$ 为 $X,Y$ 的连续或分段连续函数, 则随机变量的函数 $f(X)$ 与 $g(Y)$ 相互独立.(随机变量的函数的概念见本章第四节)

对于 $n$ 维随机变量 $(X_1,X_2,...,X_n)$, 若对任意实数 $x_1,x_2,...,x_n$, 有

$$P(X_1\le x_1,X_2\le x_2,...,X_n\le x_n)=P(X_1\le x_1)P(X_2\le x_2)...P(X_n\le x_n),$$

即

$$F(x_1,x_2,...,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)...F_{X_n}(x_n),$$

则称 $X_1,X_2,...,X_n$ 相互独立.

## 3.2 二维离散型随机变量

### 3.2.1 定义与性质

**定义 3.4** 若二维随机变量 $(X,Y)$ 的可能取值是有限多个或可列无限多个, 则称 $(X,Y)$ 为离散型随机变量.

**定义 3.5(离散型随机变量联合分布律)** 设二维离散型随机变量 $(X,Y)$ 的所有可能取值为 $(x_i,y_j),i,j=1,2,...$ 则称

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

为离散型随机变量 $(X,Y)$ 的联合分布律.

我们可以用表格形式表示 $(X,Y)$ 的联合分布律:

| $X/Y$ | $y_1$ | $y_2$ | ... | $y_j$ | ... |
|---|---|---|---|---|---|
| $x_1$ | $p_{11}$ | $p_{12}$ | ... | $p_{1j}$ | ... |
| $x_2$ | $p_{21}$ | $p_{22}$ | ... | $p_{2j}$ | ... |
| ... | ... | ... | | ... | |
| $x_i$ | $p_{i1}$ | $p_{i2}$ | ... | $p_{ij}$ | ... |
| ... | ... | ... | | ... | |

表 5

离散型随机变量的联合分布律具有如下性质:

(1) $p_{ij}\ge 0,i,j=1,2,...;$

(2) $\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty}p_{ij}=1.$

### 3.2.2 边缘分布及独立性条件

在二维联合分布中, 可推导单个随机变量 $X$ 或 $Y$ 的分布律, 称为边缘分布律.

$$P(X=x_i)=P(X=x_i,Y<+\infty)=P\left(X=x_i,\bigcup_{j=1}^{\infty}(Y=y_j)\right)$$
$$=\sum_{j=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{j=1}^{\infty}p_{ij}\triangleq p(i\cdot),$$

同理

$$P(Y=y_j)=\sum_{i=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{i=1}^{\infty}p_{ij}\triangleq p(\cdot j).$$

边缘分布律满足一维随机变量分布律的性质.

二维离散型随机变量 $(X,Y)$ 的联合分布函数为

$$F(x,y)=\sum_{x_i\le x}\sum_{y_i\le y}p_{ij},$$

其中, 和式是对一切满足 $x_i\le x,y_j\le y$ 的 $i,j$ 求和.

边缘分布函数为

$$F_X(x)=\sum_{x_i\le x}^{\infty}p_{i\cdot}=\sum_{x_i\le x}\sum_{j=1}^{\infty}p_{ij},\quad F_Y(y)=\sum_{y_j\le y}p_{\cdot j}=\sum_{y_j\le y}\sum_{i=1}^{\infty}p_{ij}.$$

二维离散型随机变量 $(X,Y)$ 在区域 $D$ 取值的概率可以表示为

$$P((X,Y)\in D)=\sum_{x_i,y_j\in D}p_{ij}.$$

二维离散型随机变量 $(X,Y)$ 的独立性等价于: 对任意 $i,j=1,2,...$ 随机事件 $\{X=x_i\}$ 与 $\{Y=y_j\}$ 独立, 即

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j),i,j=1,2,...,$$

即

$$p_{ij}=p_{i\cdot}p_{\cdot j},\quad i,j=1,2,...$$

### 3.2.3 常用二维离散型随机变量

**1. 三项分布**

若二维离散型随机变量 $(X,Y)$ 的分布律为

$$P(X=i,Y=j)=\frac{n!}{i!y!(n-i-j)!}p_1^ip_2^j(1-p_1-p_2)^{n-i-j},$$
$$i,j=0,1,...,n,i+j\le n.$$

其中,$0 < p_1, p_2 < 1, p_1 + p_2 < 1$, 则称 $(X, Y)$ 服从参数为 $n, p_1, p_2$ 的三项分布, 记为 $T(n, p_1, p_2)$. 三项分布是二项分布的推广, 是多项分布的一种.

在 $n$ 重独立试验中, 若每次试验有三种结果,$A_1, A_2, A_3$,$P(A_i) = p_i, p_1 + p_2 + p_3 = 1$, 设 $A_1$ 发生 $X$ 次,$A_2$ 发生 $Y$ 次, 则 $(X, Y)$ 服从三项分布.$P(X = i, Y = j)$ 表示 $A_1$ 发生 $i$ 次, $A_2$ 发生 $j$ 次, $A_3$ 发生 $n - i - j$ 次的概率.

**例 3.1** 证明三项分布的边缘分布是二项分布, 即若 $(X, Y) \sim T(n, p_1, p_2)$, 则 $X \sim B(n, p_1), Y \sim B(n, p_2)$.

**证明** 以 $X$ 的边缘分布为例, 对 $k = 1, 2, ..., n$

$$\begin{aligned}
P(X = k) &= \sum_{j=0}^{n-k} P(X = k, Y = j) \\
&= \sum_{j=0}^{n-k} \frac{n!}{k!j!(n-k-j)!} p_1^k p_2^j (1 - p_1 - p_2)^{n-k-j} \\
&= \frac{n!}{k!(n-k)!} p_1^k \sum_{j=0}^{n-k} \frac{(n-k)!}{j![(n-k)-j]!} p_2^j (1 - p_1 - p_2)^{(n-k)-j} \\
&= \frac{n!}{k!(n-k)!} p_1^k (1 - p_1)^{n-k} = C_n^k p_1^k (1 - p_1)^{n-k}.
\end{aligned}$$

可见, $X$ 服从二项分布 $B(n, p_1)$.

**2. 二维超几何分布**

若二维离散型随机变量 $(X, Y)$ 的分布律为

$$P(X = n_1, Y = n_2) = \frac{C_{N_1}^{n_1} C_{N_2}^{n_2} C_{N_3}^{n_3}}{C_N^n},$$

其中, $N = N_1 + N_2 + N_3, n = n_1 + n_2 + n_3$, 则称 $(X, Y)$ 服从二维超几何分布.

总共 $N$ 个元素分为 $A, B, C$ 三类, 每类分别有 $N_1, N_2, N_3$ 个, 从中不放回地取出 $n$ 个, 设取到 $A$ 类 $X$ 个, $B$ 类 $Y$ 个, 则 $(X, Y)$ 服从二维超几何分布. $P(X = n_1, Y = n_2)$ 表示取到 $A$ 类 $n_1$ 个, $B$ 类 $n_2$ 个, $C$ 类 $n_3$ 个的概率.

## 3.3 二维连续型随机变量

### 3.3.1 定义与性质

**定义 3.6** 对于二维随机变量 $(X, Y)$ 的分布函数 $F(x, y)$, 如果存在非负函数 $p(x, y)$, 使得对于任意的 $(x, y)$ 有

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

则称 $(X, Y)$ 是二维连续型随机变量, 称 $p(x, y)$ 为 $(X, Y)$ 的联合概率密度函数. 简称概率密度.

二维连续型随机变量的联合概率密度具有下列性质:

(1) $p(x, y) \geq 0$;

(2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) \mathrm{d}x \mathrm{d}y = F(+\infty, +\infty) = 1$;

(3) 设 $D$ 是平面区域, 则随机点 $(X,Y)$ 落入 $D$ 内的概率为

$$P((X,Y)\in D)=\iint_D p(x,y)\mathrm{d}x\mathrm{d}y;$$

(4) 若 $p(x,y)$ 在点 $(x,y)$ 处连续, 则有

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

由偏导数和积分性质, 当 $\Delta x,\Delta y$ 充分小时,

$$\begin{aligned}&P(x<X\leq x+\Delta x,y<Y\leq y+\Delta y)\\&=\int_x^{x+\Delta x}\int_y^{y+\Delta y}p(u,v)\mathrm{d}u\mathrm{d}v\approx p(x,y)\Delta x\Delta y.\end{aligned}$$

因此, $p(x,y)$ 的大小反映了 $(X,Y)$ 落在点 $(x,y)$ 附近的概率大小.

### 3.3.2 边缘密度及独立性条件

在研究二维随机变量 $(X,Y)$ 时, $X$ 和 $Y$ 作为一维随机变量, 其概率密度称为边缘密度, 分别记为 $p_X(x)$, $p_Y(y)$. 以下讨论边缘密度与联合密度的关系, 设 $X$ 和 $Y$ 的边缘分布, 函数分别为 $F_X(x),F_Y(y)$, 则

$$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}p(t,y)\mathrm{d}t\mathrm{d}y=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty}p(t,y)\mathrm{d}y\right]\mathrm{d}t,$$

因此,X 的边缘密度函数

$$p_X(x)=F_X'(x)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y,$$

同理,Y 的边缘密度函数

$$p_Y(y)=F_Y'(y)=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}x.$$

下面我们引入用密度函数表示的独立性条件. 随机变量 $X,Y$ 的独立性条件为, 联合密度函数可以表示为边缘密度函数的乘积:

$$F(x,y)=F_X(x)F_Y(y),\forall(x,y),$$

即

$$\begin{aligned}\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)\mathrm{d}u\mathrm{d}v&=\int_{-\infty}^{x}p_X(t)\mathrm{d}t\int_{-\infty}^{y}p_Y(t)\mathrm{d}t\\&=\int_{-\infty}^{x}\int_{-\infty}^{y}p_X(x)p_Y(y)\mathrm{d}x\mathrm{d}y.\end{aligned}$$

上式对任意 $(x,y)$ 成立, 因此得到用密度函数表示的独立性条件

$$p(x,y)=p_X(x)p_Y(y),\forall(x,y).$$

以上结果可以推广到 $n$ 维随机变量 $(X_1, X_2, ..., X_n)$.

**定义 3.7** 设 $(X_1, X_2, ..., X_n)$ 为 $n$ 维随机变量, 若存在非负函数 $p(x_1, x_2, ..., x_n)$ 使得对任意实数 $x_1, x_2, ..., x_n$, 有

$$F(x_1, x_2, ..., x_n) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} p(u_1, u_2, ..., u_n)\mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_n,$$

则称 $(X_1, X_2, ..., X_n)$ 为 $n$ 维连续型随机变量, $p(x_1, x_2, ..., x_n)$ 为 $(X_1, X_2, ..., X_n)$ 的联合密度函数. $n$ 维密度函数的性质与二维密度函数相似, 不再一一列出.

与二维情况类似, 我们可以求得 $(X_1, X_2, ..., X_n)$ 的 $k$ 维边缘密度函数 $(1 \le k < n)$, 例如, $(X_1, X_2, ..., X_n)$ 中 $X_1$ 的边缘密度函数为

$$p_{X_1}(x_1) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} p(x_1, u_2, ..., u_n)\mathrm{d}u_2\mathrm{d}u_3\cdots\mathrm{d}u_n,$$

$(X_1, X_2, ..., X_n)$ 中 $(X_1, X_2)$ 的边缘密度函数为

$$p_{(X_1,X_2)}(x_1, x_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} p(x_1, x_2, ..., u_n)\mathrm{d}u_3\mathrm{d}u_4\cdots\mathrm{d}u_n,$$

$n$ 维随机变量的独立性条件可以表示为: 若对任意实数 $x_1, x_2, ..., x_n$, 有

$$p(x_1, x_2, ..., x_n) = p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n),$$

则 $(X_1, X_2, ..., X_n)$ 相互独立.

**例 3.2** 设二维随机变量 $(X, Y)$ 的密度函数为

$$p(x,y) = \begin{cases} 1/2, (x,y) \in D \\ 0, \text{其他}. \end{cases}$$

其中, 区域 $D$ 由 $\left\{y = \dfrac{1}{x}, y = 0, x = 1, x = e^2\right\}$ 围成 (图 13), 求 $X, Y$ 的边缘密度函数.

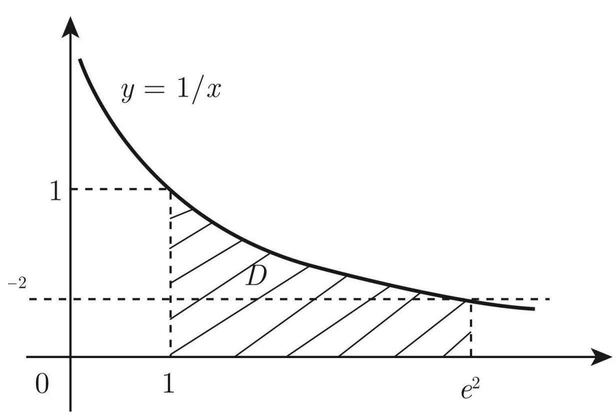


图 13

解

$$
\begin{aligned}
p_X(x) &= \int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}y \\
&= \begin{cases} \displaystyle\int_0^{1/x} \frac{1}{2}\,\mathrm{d}y & 1 \le x \le e^2 \\ 0 & \text{其他.} \end{cases} \\
&= \begin{cases} \dfrac{1}{2x} & 1 \le x \le e^2 \\ 0 & \text{其他.} \end{cases}
\end{aligned}
$$

$$
p_Y(y) = \int_{-\infty}^{+\infty} p(x,y)\mathrm{d}x = \begin{cases} \displaystyle\int_1^{e^2} \frac{1}{2}\mathrm{d}x = \frac{1}{2}(e^2-1), 0 \le y \le e^{-2} \\ \displaystyle\int_1^{1/y} \frac{1}{2}\mathrm{d}x = \frac{1}{2}(\frac{1}{y}-1), e^{-2} \le y \le 1 \\ 0, \text{其他.} \end{cases}
$$

### 3.3.3 常用二维连续型随机变量

**1．二维均匀分布**

设 $D$ 为平面有界区域, 其面积为 $S_D$, 若二维连续型随机变量 $(X,Y)$ 的联合密度为

$$
p(x,y) = \begin{cases} 1/S_D, (x,y) \in D \\ 0, \text{其他.} \end{cases}
$$

则称 $(X,Y)$ 服从区域 $D$ 上的二维均匀分布.

二维均匀分布的特点: 随机点落在 $D$ 内任一子区域 $D_k$ 的概率与 $D_k$ 的面积成正比, 与 $D_k$ 的形状和 $D_k$ 在 $D$ 内的位置无关, 这是因为

$$
P((X,Y) \in D_k) = \iint_{D_k} p(x,y)\mathrm{d}x\mathrm{d}y = \iint_{D_k} 1/S_D\mathrm{d}x\mathrm{d}y = S_{D_k}/S_D.
$$

由此可见, 利用二维均匀分布所求概率结果为面积之比, 这正是平面上的几何概型求概率的方法.

**2．二维正态分布**

若二维随机变量 $(X,Y)$ 的联合密度为

$$
p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[ \left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 \right] \right\}.
$$

其中,$\mu_1, \mu_2, \sigma_1 > 0, \sigma_2 > 0, |\ \rho\ | < 1$ 均为常数, 则称 $(X,Y)$ 服从二维正态分布, 记为 $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

一般地, $n$ 维正态分布的密度函数如下:

$$
p(x_1, x_2, ..., x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}, \quad (x_1, x_2, ..., x_n) \in \mathbb{R}^n
$$

其中 $\mu=(\mu_1,...,\mu_n)$,$\Sigma=(\sigma_{ij})$ 为正定阵. 记为 $N_n(\mu,\Sigma)$.
特别地, 在二维正态分布中

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},\Sigma^{-1}=\frac{1}{(1-\rho^2)\sigma_1^2\sigma_2^2}\begin{pmatrix}\sigma_2^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}.$$

下面给出二维正态分布的两个重要性质:

**定理 3.1** 设 $(X,Y)$ 服从二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则 X,Y 的边缘分布为 $X\sim N(\mu_1,\sigma_1^2),\quad Y\sim N(\mu_2,\sigma_2^2)$.

**证明** 令 $u=\dfrac{x-\mu_1}{\sigma_1},v=\dfrac{y-\mu_2}{\sigma_2}$, 则 $(X,Y)$ 的联合密度

$$\begin{aligned}p(x,y)&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\\&\quad\cdot\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right]\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}[\rho^2u^2+v^2-2\rho uv+(1-\rho^2)u^2]\right\}\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{(v-\rho u)^2}{2(1-\rho^2)}-\frac{u^2}{2}\right].\end{aligned}$$

$X$ 的边缘密度为

$$\begin{aligned}p_X(x)&=\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y\\&=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{u^2}{2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left[-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right]\mathrm{d}v.\end{aligned}$$

因为积分中的被积函数是 $N(\rho u,1-\rho^2)$ 的密度函数, 所以积分值为 1, 从而,

$$p_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{u^2}{2}}=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},$$

即 $X\sim N(\mu_1,\sigma_1^2)$, 同理 $Y\sim N(\mu_2,\sigma_2^2)$.
注: 配方法配凑时若配前两项则对 $u$ 做积分, 求的是 $Y$ 的边缘密度, 反之则反.
**定理 3.2** 设 $(X,Y)$ 服从二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则 $X,Y$ 相互独立的充要条件为 $\rho=0$.

**证明** 充分性: 设 $\rho=0$, 二维正态分布 $(X,Y)$ 的密度函数为

$$\begin{aligned}p(x,y)&=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}\\&=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}\cdot\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}\\&=p_X(x)p_Y(y).\end{aligned}$$

所以,$X$ 与 $Y$ 相互独立.
必要性: 设 $X,Y$ 相互独立, 则对任意实数 $x,y$ 有

$$p(x,y)=p_X(x)\cdot_Y(y),$$

特别地, 令 $x=\mu_1,y=\mu_2$ 由 $p(\mu_1,\mu_2)=p_X(\mu_1)\cdot p_Y(\mu_2)$ 得到:

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

从而得到 $\rho=0$.

## 3.4 * 条件分布

在随机事件的研究中, 讨论了条件概率问题, 而在随机变量的研究中, 可以类似讨论条件分布问题.

我们先看一维随机变量的条件分布问题, 讨论随机事件 $A=\{X\in S\}$ 发生条件下的条件分布函数. 设 $X$ 的分布函数为

$$F(x)=P(X\leq x),$$

则 $A$ 发生条件下 $X$ 的条件分布函数为

$$F(x\mid A)=P(X\leq x\mid A).$$

对于二维随机变量 $(X,Y)$, 我们分别按照离散型和连续型随机变量的场合, 讨论在 $X$ 或 $Y$ 取定某个值的条件下, 另一随机变量的条件分布.

### 3.4.1 离散型随机变量的条件分布

设 $(X,Y)$ 为二维离散型随机变量, 其概率分布律为

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

当 $P(Y=y_j)=p_(\cdot j)>0$ 时, 可得

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

称为 $Y=y_j$ 条件下, 随机变量 $X$ 的条件分布律, 可记为

| $X\mid_{Y=y_j}$ | $x_1$ | $x_2$ | ... | $x_i$ | ... |
|---|---|---|---|---|---|
| $P$ | $p_{1j}/p_{\cdot j}$ | $p_{2j}/p_{\cdot j}$ | ... | $p_{ij}/p_{\cdot j}$ | ... |

表 6

同样, 当 $P(X=x_i)=p_{i\cdot}>0$ 时, 可得 $X=x_i$ 条件下, 随机变量 $Y$ 的条件分布律. 容易验证, 上述条件分布具有离散型随机变量分布律的一切性质.

当 $X,Y$ 相互独立时, $p_{ij}=p_{i\cdot}p{\cdot}j$, 有

$$P(X=x_i\mid Y=y_j)=\frac{p_{ij}}{p_{\cdot j}}=\frac{p_{i\cdot}p{\cdot}j}{p_{\cdot j}}=p_{i\cdot}=P(X=x_i),$$

同样

$$P(Y=y_j\mid X=x_i)=p_{\cdot j}=P(Y=y_j),$$

所以, 当 $X,Y$ 独立时, 条件分布成为无条件分布.

### 3.4.2 连续型随机变量的条件分布

对于二维连续型随机变量 $(X,Y)$, 由于 $P(Y=y)=0$, 因此不能直接用

$$P(X\le x\mid Y=y)=\frac{P(X\le x,Y=y)}{P(Y=y)}$$

来定义 $Y=y$ 时 $X$ 的条件分布函数.

**定义 3.8** 对给定的 $y$, 设 $\varepsilon>0$, 若对任意 $x$, 极限

$$\lim_{\varepsilon\to 0}P(X\le x\mid y\le Y<y+\varepsilon)=\lim_{\varepsilon\to 0}\frac{P(X\le x,y\le Y<y+\varepsilon)}{P(y\le Y<y+\varepsilon)}$$

存在, 则称为 $Y=y$ 条件下,$X$ 的条件分布函数, 记为 $F_{X|Y=y}(x)$.

若 $(X,Y)$ 的分布函数为 $F(x,y)$, 密度函数为 $p(x,y)$,$Y$ 的边缘密度为 $p_Y(y)$, 则

$$\begin{aligned}F_{X|Y}(x)&=\lim_{\varepsilon\to 0}\frac{P(X\le x,y\le Y<y+\varepsilon)}{P(y\le Y<y+\varepsilon)}=\lim_{\varepsilon\to 0}\frac{F(x,y+\varepsilon)-F(x,y)}{F_Y(y+\varepsilon)-F_Y(y)}\\&=\frac{\lim\limits_{\varepsilon\to 0}\dfrac{F(x,y+\varepsilon)-F(x,y)}{\varepsilon}}{\lim\limits_{\varepsilon\to 0}\dfrac{F_Y(y+\varepsilon)-F_Y(y)}{\varepsilon}}=\frac{\dfrac{\partial F(x,y)}{\partial y}}{\dfrac{\mathrm{d}}{\mathrm{d}y}F_Y(y)}\\&=\frac{\dfrac{\partial}{\partial y}\displaystyle\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)\mathrm{d}u\mathrm{d}v}{p_Y(y)}=\frac{\displaystyle\int_{-\infty}^{x}p(u,y)\mathrm{d}u}{p_Y(y)}=\int_{-\infty}^{x}\frac{p(u,v)}{p_Y(y)}\mathrm{d}u.\end{aligned}$$

所以,

$$F_{X|Y}=y(x)=\int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}\mathrm{d}u.$$

对 $F_{X|Y}=y(x)$ 求导, 得到 $Y=y$ 条件下, $X$ 的条件密度函数, 记为 $p_{X|Y}=y(x)$,

$$p_{X|Y}=y(x)=\frac{p(x,y)}{p_Y(y)}.$$

类似可得, 在 $X=x$ 条件下,$Y$ 的条件分布函数, 条件密度函数.
当 $X,Y$ 相互独立时, $p(x,y)=p_X(x)p_Y(y)$, 从而

$$p_{X|Y=y}(x)=\frac{p(x,y)}{p_Y(y)}=p_X(x),\quad p_{Y|X=x}(y)=\frac{p(x,y)}{p_X(x)}=p_Y(y),$$

条件密度成为无条件密度.

## 3.5 二维随机变量函数的分布

对于二维随机变量 $(X,Y)$ 及实函数 $z=g(x,y)$, 可定义随机变量 $(X,Y)$ 的函数 $Z=g(X,Y)$, 即对 $(X,Y)$ 做随机试验, 得 $(x,y)$, 以 $z=g(x,y)$ 作为随机变量 $Z$ 的取值, 从而得到随机变量 $(X,Y)$ 的函数 $Z=g(X,Y)$. 若已知 $(X,Y)$ 的概率分布, 我们将研究随机变量 $Z$ 的概率分布问题, 按照 $(X,Y)$ 是离散型或连续型分别讨论.

### 3.5.1 离散型随机变量函数的分布

设二维离散型随机变量 $(X,Y)$ 的分布律为

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

若 $(X,Y)$ 的函数 $Z=g(X,Y)$, 设 $Z$ 的可能取值为 $z_k,k=1,2,...$ 则 $Z$ 的概率分布为

$$\begin{aligned}P(Z=z_k)&=P(g(X,Y)=z_k)\\&=\sum_{g(x_i,y_j)=z_k}P(X=x_i,Y=y_j)=\sum_{g(x_i,y_j)=z_k}p_{ij}.\end{aligned}$$

**例 3.3** 设 $X$ 和 $Y$ 独立, 均服从几何分布 $P(X=k)=P(Y=k)=pq^{k-1}(k=1,2,...)$, 试求 $Z=\max(X,Y)$ 的概率分布律.

**解** $\{Z=n\}=\left\{\bigcup_{k=1}^{n}(X=n,Y=k)\right\}\cup\left\{\bigcup_{k=1}^{n-1}(X=k,Y=n)\right\}$, 由于 $X$ 和 $Y$ 独立, 所以对 $n=$

$1,2,...$ 有

$$\begin{aligned}
P(Z=n) &= P\left(\bigcup_{k=1}^{n}(X=n,Y=k)\right)+P\left\{\bigcup_{k=1}^{n-1}(X=k,Y=n)\right\} \\
&= \sum_{k=1}^{n}P(X=n,Y=k)+\sum_{k=1}^{n-1}P(X=k,Y=n) \\
&= \sum_{k=1}^{n}P(X=n)P(Y=k)+\sum_{k=1}^{n-1}P(X=k)P(Y=n) \\
&= \sum_{k=1}^{n}p^2q^{n+k-2}+\sum_{k=1}^{n-1}p^2q^{n+k-2} \\
&= p^2q^{n-1}\left(\frac{1-q^n}{1-q}+\frac{1-q^{n-1}}{1-q}\right)=pq^{n-1}(2-q^n-q^{n-1}).
\end{aligned}$$

**例 3.4** 设 $X$ 和 $Y$ 独立, 分别服从参数为 $\lambda_1$ 和 $\lambda_2$ 泊松分布, 试求 $Z=X+Y$ 的概率分布律.

**解** $X$ 与 $Y$ 的可能取值都是 $0,1,2,...$ 所以 $Z=X+Y$ 的所有可能取值也是 $0,1,2,...$ 对任意 $n=0,1,2,...,$

$$\begin{aligned}
P(Z=n) &= P(X+Y=n)=P\left(\bigcup_{k=0}^{n}(X=k,Y=n-k)\right) \\
&= \sum_{k=0}^{n}P(X=k)P(Y=n-k)=\sum_{k=0}^{n}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n.
\end{aligned}$$

所以,$Z=X+Y$ 服从参数为 $\lambda_1+\lambda_2$ 的泊松分布, 这称为泊松分布的**可加性**. 泊松分布的可加性可以推广到 $n$ 个独立泊松分布随机变量的情况, 即若随机变量 $X_1,X_2,...,X_n$ 相互独立,$X_k\sim P(\lambda_k),k=1,2,...,n$, 则

$$X_1+X_2+...+X_n\sim P(\lambda_1+\lambda_2+...+\lambda_n).$$

另外, 二项分布也具有分布可加性: 若 $X\sim B(k_1,p),Y\sim B(k_2,p)$ 且 $X,Y$ 独立, 则 $Z=X+Y\sim B(k_1+k_2,p)$.

**证明**

$$\begin{aligned}P(Z=n)=P(X+Y=n)=P\left(\bigcup_{k=0}^{n}(X=k,Y=n-k)\right)&=\sum_{k=0}^{n}P(X=k)P(Y=n-k)\\&=\sum_{k=0}^{n}C_{k_1}^{k}p^{k}(1-p)^{k_1-k}C_{k_2}^{n-k}p^{n-k}(1-p)^{k_2-(n-k)}\\&=\sum_{k=0}^{n}C_{k_1}^{k}C_{k_2}^{n-k}p^{n}(1-p)^{k_1+k_2-n}\\&=C_{k_1+k_2}^{n}p^{n}(1-p)^{k_1+k_2-n}\\&\sim B(k_1+k_2,p).\end{aligned}$$

同样, 二项分布的可加性也能推广到 $n$ 个独立随机变量的情形, 即若随机变量 $X_1,X_2,...,X_n$ 相互独立,$X_i\sim B(k_i,p),i=1,2,...,n$, 则

$$X_1+X_2+...+X_n\sim B(k_1+k_2+...+k_n,p).$$

### 3.5.2 连续型随机变量函数的分布

对于二维连续型随机变量 $(X,Y)$, 其函数 $Z=g(X,Y)$ 可能是离散型, 也可能是连续型. 对于离散型的情况, 可以按照 $Z$ 的每个取值, 根据相应 $(X,Y)$ 的取值范围, 分别求其概率, 得到 $Z$ 的分布律, 通过例子说明. **例 3.5** 设 $(X,Y)$ 服从区域 $D=\{-1\le x\le 2,-3\le y\le 1\}$ 上的均匀分布,$Z=\{k,当(X,Y)\in 第k象限,k=1,2,3,4\}$, 求 $Z$ 的分布律.

**解**

$$\begin{aligned}&P(Z=1)=P((X,Y)\in 第1象限)=2/12,\quad P(Z=2)=P((X,Y)\in 第2象限)=1/12,\\&P(Z=3)=P((X,Y)\in 第3象限)=3/12,\quad P(Z=4)=P((X,Y)\in 第4象限)=6/12,\end{aligned}$$

得到 $Z$ 的分布律

| $Z$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P$ | 1/6 | 1/12 | 1/4 | 1/2 |

表 7

对于 $Z=g(X,Y)$ 是连续型的情况, 设已知 $(X,Y)$ 的密度函数 $p(x,y)$, 求 $Z=g(X,Y)$ 的密度函数. 和第二章类似, 通常利用分布函数法来求解, 即

(1) 求 $Z$ 的分布函数 $F_Z(z)$

$$F_Z(z)=P(Z\le z)=P(g(X,Y)\le z)=P((X,Y)\in D_z)=\iint_{D_z}p(x,y)\mathrm{d}x\mathrm{d}y,$$

其中,$D_z=\{(x,y)\mid g(x,y)\le z\}$.

(2) 求 $Z$ 的密度函数:$p_Z(z)=F_Z'(z)$.

**例 3.6** 设 $X, Y$ 相互独立, $X$ 服从于 $U(0,1)$, $Y$ 服从于 $E(1)$, 求 $Z = 2X + Y$ 的概率密度函数.

**解** $<$ 定义法 $>$ $(X, Y)$ 的联合密度函数为

$$p(x,y) = p_X(x)p_Y(y) = \begin{cases} e^{-y}, 0 \le x \le 1, y \ge 0 \\ 0, \text{其他}. \end{cases}$$

$$\begin{aligned} F_Z(z) = P(Z \le z) = P(2X + Y \le z) = \iint_{2x+y \le z} p(x,y)\mathrm{d}x\mathrm{d}y. \\ z < 0, F_Z(z) = \iint_{2x+y \le z} 0\mathrm{d}x\mathrm{d}y = 0, \\ 0 \le z < 2, F_Z(z) = \int_0^{z/2} \mathrm{d}x \int_0^{z-2x} e^{(} - y)\mathrm{d}y = \frac{z}{2} + \frac{1}{2}e^{-z} - \frac{1}{2}, \\ z \ge 2, F_Z(z) = \int_0^1 \mathrm{d}x \int_0^{z-2x} e^{(} - y)\mathrm{d}y = 1 - \frac{1}{2}(e^2 - 1)e^{(} - z), \end{aligned}$$

$<$ 卷积公式法 $>$ 由 $X \sim U[0,1]$, 有 $2X \sim U[0,2]$

$$p_{2X}(x) = \begin{cases} \frac{1}{2}, 0 \le x \le 2 \\ 0, \text{其他}. \end{cases} \quad p_Y(y) = \begin{cases} e^{-y}, y \ge 0 \\ 0, \text{其他}. \end{cases}$$

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{2X}(x)p_Y(z-x)\mathrm{d}x = \begin{cases} \int_{-\infty}^{+\infty} \frac{1}{2}p_Y(z-x)\mathrm{d}x, z > 0 \\ 0, z \le 0. \end{cases}$$

故当 $0 \le z \le 2$ 时, 有 $p(z) = \int_0^z \frac{1}{2}e^{-(z-x)}\mathrm{d}x = \frac{1}{2}(1 - e^{-z})$.

当 $z > 2$ 时, 有 $p(z) = \int_0^2 \frac{1}{2}e^{-(z-x)}\mathrm{d}x = \frac{1}{2}(e^{2-z} - e^{-z})$.

$$p_Z(z) = F_Z'(z) = \begin{cases} \frac{1}{2} - \frac{1}{2}e^{-z}, 0 \le z \le 2 \\ \frac{1}{2}(e^2 - 1)e^{-z}, z \ge 2 \\ 0, z < 0. \end{cases}$$

注: 由于 $z - x$ 为 $p_Y(\cdot)$ 的自变量, 故 $z - x > 0$ 时才有效, 先 $z$ 不动, 做积分; 后讨论 $Z$, 分段.

**例 3.7** 在区间 $[0,1]$ 随机投两点, 求两点间距离 $Z$ 的分布函数与密度函数.

**解** 设两点坐标 $(X, Y)$, 则 $Z =\mid X - Y \mid$ ,$(X, Y)$ 服从 $[0,1]\times[0,1]$ 上的二维均匀分布, 其密度为

$$p(x,y) = \begin{cases} 1, 0 \le x, y \le 1 \\ 0, \text{其他}. \end{cases}$$

先求 $Z$ 的分布函数 $F_Z(z)$,

$$F_Z(z) = P(Z \leq z) = P(\mid X - Y \mid \leq z) = \iint_{|x-y|\leq z} p(x,y)\mathrm{d}x\mathrm{d}y.$$

显然, $z < 0$ 时,$F_Z(z) = 0$; $z > 1$ 时, $F_Z(z) = 1$; $0 \leq z \leq 1$ 时,

$$\begin{aligned} F_Z(z) &= \iint_{|x-y|\leq z} p(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{|x-y|\leq z} \mathrm{d}x\mathrm{d}y = S_{D_z} = 1-(1-z)^2 \\ &= 2z - z^2. \end{aligned}$$

得到

$$F_Z(z) = \begin{cases} 0, z < 0 \\ 2z - z^2, 0 \leq z \leq 1 \\ 1, z \geq 1. \end{cases}$$

$Z$ 的密度函数为

$$p_Z(z) = F'_Z(z) = \begin{cases} 2 - 2z, 0 \leq z \leq 1 \\ 0, 其他. \end{cases}$$

以下我们给出几个常用函数的分布公式.

**1．和与差的分布**

设 $(X,Y)$ 为二维随机变量, 其联合密度函数为 $p(x,y)$, 求 $Z = X + Y$ 的密度函数 $p_Z(z)$, 用分布函数法,

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \iint_{D_z} p(x,y)\mathrm{d}x\mathrm{d}y,$$

其中,$D_z = \{(x,y) \mid x + y \leq z\}$, 按 $x$ 型区域积分,

$$F_Z(z) = \int_{-\infty}^{+\infty} \mathrm{d}x \int_{-\infty}^{z-x} p(x,y)\mathrm{d}y,$$

作变量变换 $u = y + x$,

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{z} p(x, u-x)\mathrm{d}u \right) \mathrm{d}x \\ &= \int_{-\infty}^{z} \left( \int_{-\infty}^{+\infty} p(x, u-x)\mathrm{d}x \right) \mathrm{d}u, \end{aligned}$$

求导得,$p_Z(z) = F'_Z(z) = \displaystyle\int_{-\infty}^{+\infty} p(x, z-x)\mathrm{d}x$, 如果按 $y$ 型区域积分, 可得

$$p_Z(z) = \int_{-\infty}^{+\infty} p(z-y, y)\mathrm{d}y.$$

特别地, 若 $X, Y$ 相互独立时, 可以化为

$$p_Z(z)=\int_{-\infty}^{+\infty} p_X(x)p_Y(z-x)\mathrm{d}x,$$

$$p_Z(z)=\int_{-\infty}^{+\infty} p_X(z-y)p_Y(y)\mathrm{d}y,$$

以上两式称为**卷积公式**.

类似可以得到差的分布公式, 设 $Z=X-Y$ 的密度函数 $p_Z(z)$, 则

$$p_z(z)=\int_{-\infty}^{+\infty} p(x,x-z)\mathrm{d}x=\int_{-\infty}^{+\infty} p(z+y,y)\mathrm{d}y.$$

**例 3.8** 设随机变量 $X, Y$ 相互独立且有相同分布, 其密度函数均为

$$p(x)=\begin{cases} e^{1-x}, x\geq 1 \\ 0, 其他. \end{cases}$$

用卷积公式求 $Z=X+Y$ 的密度函数.

**解**　由卷积公式,$Z=X+Y$ 的密度函数

$$p_Z(z)=\int_{-\infty}^{+\infty} p_X(x)p_Y(z-x)\mathrm{d}x.$$

显然,$z$ 的可能取值范围是 $z>2$ , 所以 $z\leq 2$ 时,$p_Z(z)=0$.

由 $X$ 与 $Y$ 的密度函数可知, 仅当 $x>1$ 且 $z-x>1$ , 即 $1<x<z-1$ 时, 上述积分的被积函数不等于 0, 于是

$$p(x)=\begin{cases} \displaystyle\int_1^{z-1} e^{1-x}e^{1-(z-x)}\mathrm{d}x, z>2 \\ 0, 其他. \end{cases} \qquad p(x)=\begin{cases} (z-2)e^{2-z}, z\geq 2 \\ 0, 其他. \end{cases}$$

**例 3.9** 设随机变量 $X, Y$ 相互独立且均服从标准正态分布, 求 $Z=X+Y$ 的密度函数. **解**　$X, Y$ 的密度函数均为

$$p(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}, -\infty<t<+\infty.$$

由卷积公式

$$\begin{aligned} p_Z(z)&=\int_{-\infty}^{+\infty} p_X(x)p_Y(z-x)\mathrm{d}x \\ &=\frac{1}{2}\pi\int_{-\infty}^{+\infty} e^{-z^2/2}\cdot e^{-(z-x)^2/2}\mathrm{d}x=\frac{1}{2\pi}e^{-z^2/4}\int_{-\infty}^{+\infty} e^{-(x-z/2)^2}\mathrm{d}x \\ &=\frac{1}{2\sqrt{\pi}}e^{-x^2/4}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{\pi}}e^{-(x-z/2)^2}\mathrm{d}x=\frac{1}{2\sqrt{\pi}}e^{-z^2/4}, \end{aligned}$$

最后一个积分的被积函数是正态分布 $N(z/2,1/2)$ 的密度函数, 所以积分值为 1. 由结果可见,$Z\sim N(0,2)$.

一般情形, 可以推出, 若 $X,Y$ 相互独立,$X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$, 则 $X\pm Y\sim N(\mu_1\pm\mu_2,\sigma_1^2+\sigma_2^2)$.

更一般的情形可以推广到 $n$ 维, 设 $X_1,X_2,...,X_n$ 相互独立, $X_i\sim N(\mu_i,\sigma_i^2)$, $i=1,2,...,n$ 则 $X_1+X_2+...+X_n\sim N(\mu_1+\mu_2+...+\mu_n,\sigma_1^2+\sigma_2^2+...+\sigma_n^2)$.

下面我们不加证明地给出积和商的公式.

**2. * 积与商的分布**

设 $Z=XY$ 的密度函数 $p_Z(z)$ , 则

$$p_Z(z)=\int_{-\infty}^{+\infty}p\left(x,\frac{z}{x}\right)\frac{1}{\mid x\mid}\mathrm{d}x=\int_{-\infty}^{+\infty}p\left(\frac{z}{y},y\right)\frac{1}{\mid y\mid}\mathrm{d}y.$$

设 $Z=X/Y$ 的密度函数 $p_Z(z)$ , 则

$$p_Z(z)=\int_{-\infty}^{+\infty}p\left(z,zx\right)\mid x\mid\mathrm{d}x=\int_{-\infty}^{+\infty}p\left(zy,y\right)\mid y\mid\mathrm{d}y.$$

**3. max 与 min 的分布**

设随机变量 $X,Y$ 相互独立, 其分布函数分别为 $F_X(x)$ 和 $F_Y(y)$ , 令 $M=\max(X,Y),N=\min(X,Y)$, 以下我们推导 $M$ 和 $N$ 的分布函数.

$M=\max(X,Y)$, 对于任意 $z,M\le z$ 等价于 $X\le z$ 且 $Y\le z$, 因此

$$\begin{aligned}F_M(z)&=P(M\le z)=P(\max(X,Y)\le z)=P(X\le z,Y\le z)\\&=P(X\le z)P(Y\le z)=F_X(z)F_Y(z).\end{aligned}$$

类似有 $N=\min(X,Y)$, 对任意 $z,N>z$ 等价于 $X>z$ 且 $Y>z$,

$$\begin{aligned}F_N(z)&=P(N\le z)=P(\min(X,Y)\le z)=1-P(\min(X,Y)>z)\\&=1-P(X>z,Y>z)=1-P(X>z)P(Y>z)\\&=1-[1-F_X(z)][1-F_Y(z)].\end{aligned}$$

以上结果可以推广到 $n$ 个相互独立随机变量 $X_1,X_2,...,X_n$ 的情形, 设 $X_i$ 的分布函数为 $F_{X_i}(x_i),i=1,2,...,n$, 令 $M=\max(X_1,X_2,...,X_n)$, $N=\min(X_1,X_2,...,X_n)$, 则 $M$ 的分布函数 $F_M(z)$, $N$ 的分布函数 $F_N(z)$ 分别为

$$\begin{aligned}&F_M(z)=F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z),\\&F_N(z)=1-[1-F_{X_1}(z)][1-F_{X_2}(z)]\cdots[1-F_{X_n}(z)].\end{aligned}$$

特别地, 当随机变量 $X_1, X_2, ..., X_n$ 相互独立且具有相同的分布函数 $F(x)$ 时, 有

$$
\begin{aligned}
F_M(z) &= [F(z)]^n, \\
F_N(z) &= 1 - [1 - F(z)]^n.
\end{aligned}
$$

**例 3.10** 二维随机变量 $(X, Y)$ 在区域 $D = (x, y) : 0 < x, y < a$ 上服从均匀分布, 求 $M = \max(X, Y)$ 的概率密度函数.

**解** $(X, Y)$ 的联合密度函数为

$$
p(x, y) = \begin{cases} 1/a, 0 < x, y < 1 \\ 0, \text{其他}. \end{cases}
$$

可以求得

$$
p_X(x) = \int_{-\infty}^{+\infty} p(x, y)\,\mathrm{d}y = \begin{cases} 1/a, 0 \le x \le a \\ 0, \text{其他}. \end{cases}
$$

$$
p_Y(y) = \int_{-\infty}^{+\infty} p(x, y)\,\mathrm{d}x = \begin{cases} 1/a, 0 \le x \le a \\ 0, \text{其他}. \end{cases}
$$

可见, $p(x, y) = p_X(x)p_Y(y)$ , 所以 $X, Y$ 相互独立, 均服从 $(0, a)$ 上的均匀分布, $M = \max(X, Y)$ 的分布函数

$$
F_M(z) = F_X(z)F_Y(z) = \begin{cases} z^2/a^2, 0 < z < a \\ 0, \text{其他}. \end{cases}
$$

从而得到 $M = \max(X, Y)$ 的密度函数为

$$
p_M(z) = F_M'(z) = \begin{cases} 2z/a^2, 0 < z < a \\ 0, \text{其他}. \end{cases}
$$

以上讨论随机变量 $X, Y$ 函数的分布时,$X, Y$ 同为离散型或同为连续型. 若出现 $X, Y$ 一个是连续型, 另一个是离散型的情形, 可以用分布函数法, 把离散型随机变量的所有取值看成样本空间的划分, 由全概率公式求解.

**例 3.23** 设随机变量 $X, Y$ 独立, 其中, $P(X = k) = 1/3, k = 1, 2, 3$, $Y$ 服从指数分布 $E(1)$, 求随机变量 $Z = X + Y$ 的概率密度函数.

**解** $Y$ 的密度函数 $p_Y(y) = \begin{cases} e^{-y}, y \geq 0 \\ 0, 其他. \end{cases}$, 设 $Z = X + Y$ 的分布函数为 $F_Z(z)$, 则

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) \\ &= \sum_{k=1}^{3} P(X = k) P(X + Y \leq z \mid X = k) \\ &= \frac{1}{3} \sum_{k=1}^{3} P(Y \leq z - k) = \frac{1}{3} \sum_{k=1}^{3} F_Y(z - k), \end{aligned}$$

$$\begin{aligned} p_Z(z) = F_Z' &= \frac{1}{3}[p_Y(z-1) + p_Y(z-2) + p_Y(z-3)] \\ &= \begin{cases} 0, z < 1 \\ e^{1-z}/3, 1 \leq z < 2 \\ (e^{1-z} + e^{2-z})/3, 2 \leq z < 3 \\ (e^{1-z} + e^{2-z} + e^{3-z})/3, z \geq 3 \end{cases} \end{aligned}$$

# 4 随机变量的数字特征

我们知道随机变量或随机向量的分布函数完全刻画了随机变量或随机向量的概率特征. 然而能否用更为简单的数字来刻画随机变量的概率特征呢? 但这时就不能完整地描述随机变量的概率特征了, 只能刻画随机变量某一方面的概率性质.

例如, 考察一批灯泡的寿命, 这可以用分布函数来描述. 如果不知道分布函数, 而仅知道灯泡的平均寿命, 虽不能对灯泡的寿命状况提供一个完整的刻画, 但也给我们提供了一个重要的参考数据. 这个重要的数字特征称为数学期望, 简称为期望或均值.

又如, 对一射手进行评定时, 除了考察其射击命中的平均值, 还要了解其命中点是比较分散还是比较集中. 这个重要的数字特征就是方差. 它可用来刻画随机变量的分散程度.

有时, 我们需要比较两个随机变量之间的相关程度. 比如, 在学习过程中, 有些课程是有着密切联系的, 如数学与物理的学习, 于是可以推断一个学生的数学成绩与物理成绩之间会有比较强的相关程度. 这种描述随机变量之间相关程度的重要数字特征称为相关系数. 总之, 这些刻画随机变量性质的数字就称为随机变量的数字特征. 虽然这些数字特征不能完全地刻画随机变量, 但它们简单方便, 在实际中有着重要的应用价值.

## 4.1 数学期望

### 4.1.1 离散型随机变量的数学期望

简单地说, 数学期望就是随机变量的加权平均取值, 该概念由数学家惠更斯于 1657 年提出.

**定义 4.1** 设离散型随机变量 $X$ 的分布律为 $P(X=x_i)=p_i, i=1,2,...$ 若级数 $\sum\limits_{i=1}^{+\infty} |\, x_i \,|\, p_i$ 收敛, 则称 $\sum\limits_{i=1}^{+\infty} x_i \mid p_i$ 为 $X$ 的**数学期望**, 记为 $EX$. 即

$$EX=\sum_{i=1}^{+\infty} x_i p_i.$$

随机变量的数学期望可简称为**期望**. 数学期望 $EX$ 反映了随机变量 $X$ 取值的平均, 所以也称为**均值**. 由于数学期望是由分布函数唯一确定的, 所以也称为相应分布的数学期望. 若 $\sum\limits_{i=1}^{+\infty} |\, x_i \,|\, p_i$ 发散, 则称随机变量 $X$ 的**数学期望不存在**. 显然, 随机变量的数学期望 $EX$ 不一定存在.

事实上, 由于数学期望刻画随机变量 $X$ 取值的平均大小, 这是不应该受到随机变量取值次序的不同而改变的. 即相应的级数不论求和次序如何改变, 其值都应该保持不变. 这在数学上就相当于要求级数 $\sum\limits_{i=1}^{+\infty} |\, x_i \,|\, p_i$ 收敛, 或者说, 我们需要级数 $\sum\limits_{i=1}^{+\infty} x_i p_i$ 绝对收敛, 仅要求级数 $\sum\limits_{i=1}^{+\infty} x_i p_i$ 条件收敛是不够的. 下面我们来看几个例子.

**例 4.1(二项分布的数学期望)** 设随机变量 $X \sim B(n,p)$ , 求 $X$ 的数学期望.

**解** 二项分布的概率分布如下:

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

根据定义 4.1, 计算如下:

$$EX = \sum_{k=0}^{n} kp_k = \sum_{k=0}^{n} kC_n^k p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{k \cdot n!}{k!(n-k)!} p^k (1-p)$$
$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{(n-1)-(k-1)} = np.$$

**例 4.2(泊松分布的数学期望)** 设随机变量 $X$ 服从参数为 $\lambda$ 的泊松分布 $P(\lambda), \lambda > 0$ . 求 $X$ 的数学期望.

**解** 泊松分布的概率分布如下:

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

则 $X$ 的数学期望为

$$EX = \sum_{k=0}^{+\infty} kP(X=k) = \sum_{k=1}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{+\infty} \frac{\lambda^{(k-1)}}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda.$$

于是我们发现, 泊松分布中的参数 $\lambda$ 表示随机变量的均值. 例如, 当随机变量 $X$ 代表某高速公路发生的交通事故数时, $\lambda$ 就代表平均发生的交通事故数. 这样我们对泊松分布中的参数 $\lambda$ 就有了直观的理解.

### 4.1.2 连续型随机变量的数学期望

下面我们考虑连续型随机变量的情形. 定义如下:

**定义 4.2** 设连续型随机变量 $X$ 的密度函数为 $p(x)$, 若积分 $\int_{-\infty}^{+\infty} \mid x \mid p(x) \mathrm{d}x < +\infty$ , 则称 $\int_{-\infty}^{+\infty} \mid x \mid p(x) \mathrm{d}x$ 为 $X$ 的数学期望或均值. 记为 $EX$. 即

$$EX = \int_{-\infty}^{+\infty} xp(x) \mathrm{d}x.$$

事实上, 若 $X$ 是一个连续型随机变量, 则 $p(x)\Delta x$ 近似表示 $X$ 在 $(x, x+\Delta x]$ 内取值的概率. 而这里的积分类似于定义 4.1 中的求和, 于是可以看出, 公式 (4.1) 与公式 (4.2) 本质是相同的.

类似地, 连续型随机变量的数学期望也不一定存在. 例如, 连续型柯西 (Cauchy) 分布的数学期望是不存在的.

柯西分布的密度函数如下:

$$p(x) = \frac{1}{\pi(1+x^2)}, x \in (-\infty, +\infty),$$

由于

$$\int_{-\infty}^{+\infty} \frac{\mid x \mid}{\pi(1+x^2)} \mathrm{d}x = \frac{2}{\pi} \int_{0}^{+\infty} \frac{x}{1+x^2} \mathrm{d}x = +\infty,$$

所以该分布的数学期望不存在. 值得注意的是, 尽管柯西分布的密度函数是关于 $y$ 轴对称, 但它的

期望不是零, 而是不存在.

**例 4.3(均匀分布的数学期望)** 设随机变量 $X$ 服从 $[a,b]$ 上的均匀分布, 求 $X$ 的数学期望.

**解** 均匀分布的密度函数为

$$p(x)=\begin{cases}\dfrac{1}{b-a}, a\leq x\leq b\\ 0, \text{其他}.\end{cases}$$

则 $X$ 的数学期望

$$EX=\int_{-\infty}^{+\infty}xp(x)\mathrm{d}x=\int_a^b\frac{x}{b-a}\mathrm{d}x=\left.\frac{x^2}{2(b-a)}\right|_a^b=\frac{a+b}{2}.$$

**例 4.4(指数分布的数学期望)** 设随机变量 $X\sim E(\lambda),\lambda>0$ , 求 $X$ 的数学期望.

**解** 由第二章知, 指数分布 X 的密度函数为

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

于是由定义 4.2 得

$$EX=\int_{-\infty}^{+\infty}xp(x)\mathrm{d}x=\int_0^{+\infty}\lambda xe^{-\lambda x}\mathrm{d}x=-xe^{-\lambda x}\Big|_0^{+\infty}+\int_0^{+\infty}e^{-\lambda x}\mathrm{d}x=\frac{-1}{\lambda}e^{-\lambda x}\Big|_0^{+\infty}=\frac{1}{\lambda}.$$

**例 4.5(正态分布的数学期望)** 设随机变量 $X$ 服从正态分布 $N(\mu,\sigma^2)$, 求 $X$ 的数学期望.

**解** 随机变量 $X$ 的密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2}2\sigma^2, x\in\mathbb{R},$$

其数学期望的计算如下:

$$\begin{aligned}EX&=\int_{-\infty}^{+\infty}xp(x)\mathrm{d}x=\int_{-\infty}^{+\infty}\frac{x}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\mu+\sigma t)e^{\frac{-t^2}{2}}\,\mathrm{d}t=\mu\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{\frac{-t^2}{2}}\,\mathrm{d}t\\&\quad+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}te^{\frac{-t^2}{2}}\,\mathrm{d}t=\mu.\end{aligned}$$

其中, 第三个等号是运用变量变换 $t=\dfrac{x-\mu}{\sigma}$. 于是得 $X$ 的数学期望为 $\mu$ . 特别地, 标准正态分布 $N(0,1)$ 的期望为 0 . 这与正态分布的函数图像完全吻合, 因为其密度函数曲线是关于 $x=\mu$ 对称的.

### 4.1.3 随机变量函数的数学期望

随机变量的数学期望是一个经常使用的数字特征. 在实际当中, 我们还会遇到求随机变量函数的数学期望的问题.

一般地, 已知随机变量 $X$ 的分布, 且 $Y=g(X)$ 也是一个随机变量, 现需求 $Y$ 的数学期望. 若根据数学期望的定义, 我们需要求出随机变量 $Y$ 的分布. 但由上面例题的启发, 我们可以不必求出 $Y$ 的密度函数, 而直接利用 $X$ 的分布来求 $EY$. 下面我们不加证明地给出随机变量函数的期望的计算定理.

**定理 4.1** 设随机变量 $X$ 的函数 $Y=g(X)$ 也是一个随机变量, 则下面结论成立.

(1) 若 $X$ 为离散型随机变量, 其分布律为 $P(X=x_k)=p_k, k=1,2,...,$ 若 $\sum\limits_{k=1}^{+\infty} \mid g(x_k) \mid p_k$ 收敛, 则 $Y$ 的数学期望为

$$EY=E[g(X)]=\sum_{k=1}^{+\infty} g(x_k)P(X=x_k)=\sum_{k=1}^{+\infty} g(x_k)p_k.$$

(2) 若 $X$ 为连续型随机变量, 其密度函数为 $p(x)$, 若 $\int_{-\infty}^{+\infty} \mid g(x) \mid p(x)\mathrm{d}x<+\infty$, 则

$$EY=E[g(X)]=\int_{-\infty}^{+\infty} g(x)p(x)\mathrm{d}x.$$

定理 4.1 可以被推广到随机向量情形.

**定理 4.2** 设随机向量 $(X,Y)$ 的函数 $Z=g(X,Y)$ 是一个随机变量, 则下面结论成立.

(1) 若 $(X,Y)$ 为离散型的随机向量, 其联合分布律为 $P(X=x_i,Y=y_j)=p_{ij}, i,j=1,2,...,$ 若 $\sum\limits_{i=1}^{+\infty}\sum\limits_{j=1}^{+\infty} \mid g(x_i,y_j) \mid p_{ij}$ 收敛, 则

$$EZ=E[g(X,Y)]=\sum_{i=1}^{\infty}\sum_{i=1}^{\infty} g(x_i,y_j)p_{ij}.$$

(2) 若 $(X,Y)$ 为连续型的随机向量, 其联合密度为 $p(x,y)$, 若

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mid g(x,y) \mid p(x,y)\mathrm{d}x\mathrm{d}y<+\infty$$

则

$$EZ=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)p(x,y)\mathrm{d}x\mathrm{d}y.$$

这两个定理极其重要, 它们提供了计算随机变量或向量函数的期望的一般方法, 而不需要先计算随机变量或向量函数的分布. 事实上, 经过第三章的学习, 可以发现求随机变量或向量函数的分布不是一件容易的事. 但运用定理 4.1 及定理 4.2, 我们发现求随机变量函数的期望比求随机变量的分布函数要容易得多. 这也是我们使用数字特征比使用分布函数更方便的地方.

**例 4.6** 设 $X_1,X_2,...,X_n$ 均服从 $[0,1]$ 上的均匀分布, 且相互独立. 令 $X_{(n)}=\max(X_1,X_2,...,X_n)$. 试求 $EX_{(n)}$.

**解** 当随机变量 $X_1, X_2, ..., X_n$ 相互独立且具有相同的分布函数 $F(x)$ 时,$X_{(n)}$ 的分布函数为

$$[F(x)]^n = \begin{cases} 0, x < 0 \\ x^n, 0 \le x \le 1 \\ 1, x > 1. \end{cases}$$

于是 $X_{(n)}$ 的密度函数为

$$p(x) = \begin{cases} nx^{n-1}, 0 \le x \le 1 \\ 0, \text{其他}. \end{cases}$$

由定义 4.2, 得

$$EX_{(n)} = \int_{-\infty}^{+\infty} xp(x)\mathrm{d}x = \int_0^1 nx^n \mathrm{d}x = \frac{n}{n+1}.$$

本题如果仍然使用定理 4.1 进行计算, 则需要计算 n 重积分, 反而变得复杂了. 因此我们在学习过程中需要灵活运用, 不可一成不变.

### 4.1.4 数学期望的性质

下面我们来研究一下数学期望的性质, 这些性质虽然简单, 但经常会用到. 首先我们假设下列等式中所涉及的数学期望均存在.

(1) 设 $a$ 为常数, 则 $E(a) = a$.

(2) 线性性质: 对任意有限常数 $a, b, E(aX + bY) = aEX + bEY$.

**证明** 这里我们仅对连续型情况加以证明. 不妨设 $(X, Y)$ 的联合密度函数为 $p(x, y)$, 利用定理 4.2 可知

$$\begin{aligned} E(aX + bY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (ax + by)p(x, y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} axp(x, y)\mathrm{d}x\mathrm{d}y + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} byp(x, y)\mathrm{d}x\mathrm{d}y \\ &= a\int_{-\infty}^{+\infty} x\left(\int_{-\infty}^{+\infty} p(x, y)\mathrm{d}y\right)\mathrm{d}x + b\int_{-\infty}^{+\infty} y\left(\int_{-\infty}^{+\infty} p(x, y)\mathrm{d}x\right)\mathrm{d}y \\ &= a\int_{-\infty}^{+\infty} xpx(x)\mathrm{d}x + b\int_{-\infty}^{+\infty} ypy(y)\mathrm{d}y = aEX + bEY. \end{aligned}$$

一般地, 对任意 $n$ 个实数 $a_i, i = 1, 2, ..., n$ 和任意 $n$ 个期望有限的随机变量 $X_1, ..., X_n$, 有 $E(a_1X_1 + a_2X_2 + ... + a_nX_n) = a_1EX_1 + a_2EX_2 + ... + a_nEX_n$.

(3) 若 $X, Y$ 相互独立, 则 $E(XY) = EX \cdot EY$.

**证明** 同样我们仅对连续型加以证明, 假设随机变量 $X$ 和 $Y$ 的密度函数分别为 $p_X(x), p_Y(y)$. 于

是

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp_X(x)p_Y(y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} xp_X(x)\mathrm{d}x\right) yp_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} xp_X(x)\mathrm{d}x \cdot \int_{-\infty}^{+\infty} yp_Y(y)\mathrm{d}y \\
&= EX \cdot EY.
\end{aligned}
$$

一般地, 若 $X_1, ..., X_n$ 相互独立, 则

$$
E(X_1X_2...X_n) = EX_1EX_2...EX_n.
$$

在求解数学期望时, 如果能恰当利用上述性质, 将使求解过程变得更为简捷. 下面我们来看一些例题. 首先运用期望的性质计算二项分布的期望.

**例 4.7(超几何分布的数学期望)** 在 $N$ 件产品中含 $M$ 件次品, 其余为正品. 现从中任意不放回地取出 $n$ 件, 求抽得的平均次品数.

**解** 令 $X$ 为抽到的次品数. 由第一章的例 1.5 知 $P(X=k) = (C_M^kC_(N-M)^(n-k))/(C_N^n)$, 如果直接计算期望, 则计算较为复杂. 运用期望性质则可以简化运算. 首先定义

$$
X_i = \begin{cases} 1, \text{第 i 件次品被抽到} \\ 0, \text{第 i 件次品未被抽到.} \end{cases}
$$

则 $X = X_1 + X_2 + ... + X_M$, 且

$$
P(X_i = 1) = \frac{C_1^1C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N},
$$

而 $EX_i = P(X_i = 1) = n/N$. 利用性质 (2), 得

$$
EX = E(X_1 + X_2 + ... + X_M) = EX_1 + EX_2 + ... + EX_M = \frac{nM}{N}.
$$

**例 4.8** 将 $n$ 个球放入 $M$ 个盒子, 设每个球落入各个盒子是等可能的. 求有球盒子数 $X$ 的数学期望.

**解** 令随机变量

$$
X_i = \begin{cases} 1, \text{第 i 个盒子中有球} \\ 0, \text{第 i 个盒子中无球.} \end{cases}
$$

故而有 $X = X_1 + X_2 + ... + X_M$, 且 $X_i$ 的数学期望为

$$
E(X_i) = P(X_i = 1) = 1 - P(X_i = 0) = M\left[1 - \left(\frac{M-1}{M}\right)^n\right].
$$

所以 $EX = EX_1 + EX_2 + ... + EX_M = MEX_1 = M\left[1-\left(\frac{M-1}{M}\right)^n\right]$.

**例 4.9** 设随机变量 $X$ 和 $Y$ 独立, 且 $X \sim U[-1,1]$,$Y \sim E(2)$, 求 $E(X^2Y)$.

**解** 由随机变量 $X$ 和 $Y$ 独立, 得到随机变量 $X^2$ 和 $Y$ 也相互独立. 于是运用期望的性质 (3), 有

$$E(X^2Y) = E(X^2)\cdot EY = \left(\int_{-1}^{1}\frac{1}{2}x^2\mathrm{d}x\right)\times\frac{1}{2} = \frac{1}{6}.$$

## 4.2 方差

### 4.2.1 方差的定义

除了数学期望之外, 另一个重要的数字特征是方差, 它反映了随机变量的分散程度. 首先可以用 $X-EX$ 来表示随机变量相对于期望的偏差, 由于离散程度是不受正负号影响的, 于是我们采用 $(X-EX)^2$ 来刻画随机变量的离散程度. 最后对其求期望得 $E(X-EX)^2$. 我们就用这个数字来衡量随机变量的分散程度. 具体定义如下:

**定义 4.3** 设 $X$ 是一个随机变量, 若 $EX^2 < +\infty$ , 则称 $E(X-EX)^2$ 为随机变量 $X$ 的**方差**, 记 $D(X)$ 或 $\mathrm{Var}(X)$. 同时称 $\sqrt{D(X)}$ 为 $X$ 的**均方差**或**标准差**, 记为 $\sigma(X)$.

当 $X$ 为离散型随机变量, 其分布律为 $P(X=x_k)=p_k, k=1,2,...$ 时,

$$D(X) = E(X-EX)^2 = \sum_{k=1}^{+\infty}(x_k-EX)^2p_k.$$

当 X 为连续型随机变量, 其密度函数为 $p(x)$ 时,

$$D(X) = E(X-EX)^2 = \int_{-\infty}^{+\infty}(x-EX)^2p(x)\mathrm{d}x.$$

此外, 一般地, 我们还可以得到方差下面的计算公式. 由期望的线性性质得

$$\begin{aligned}D(X) &= E(X-EX)^2 = E[X^2-2XEX+(EX)^2]\\ &= EX^2-2EX\cdot EX+(EX)^2 = EX^2-(EX)^2.\end{aligned}$$

于是

$$D(X) = EX^2-(EX)^2.$$

这是计算方差的常用公式.

**例 4.10(二项分布的方差)** 设 $X \sim B(n,p)$, 试求 $D(X)$.

**解** 二项分布的概率如下:

$$P(X=k) = C_n^kp^k(1-p)^{n-k}, k=0,1,...,n,$$

其数学期望为 $EX=np$. 根据公式, 只需计算 $EX^2$, 具体如下:

$$\begin{aligned}EX^2&=\sum_{k=0}^{n}k^2C_n^kp^k(1-p)^{n-k}=\sum_{k=1}^{n}\frac{k^2\cdot n!}{k!(n-k)!}p^k(1-p)^{n-k}\\&=np\sum_{k=1}^{n}\frac{[(k-1)+1](n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{(n-1)-(k-1)},\end{aligned}$$

$$\begin{aligned}&=np\sum_{k=1}^{n}\frac{(k-1)(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{(n-1)-(k-1)}\\&+np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{(n-1)-(k-1)}\\&=np[(n-1)p+1]=(np)^2+np(1-p).\end{aligned}$$

得

$$D(X)=EX^2-(EX)^2=(np)^2+np(1-p)-(np)^2=np(1-p)=npq.$$

类似地, 设随机变量 $X$ 服从参数为 $\lambda$ 的泊松分布 $P(\lambda)$, $\lambda>0$. 则 $X$ 的方差 $D(X)=\lambda$. 注意泊松分布的期望和方差均为 $\lambda$.

**例 4.11(均匀分布的方差)** 设随机变量 $X$ 服从 $[a,b]$ 上的均匀分布, 试求 $D(X)$.

**解** 均匀分布的密度函数为

$$p(x)=\begin{cases}\dfrac{1}{b-a},a\le x\le b\\0,\text{其他}.\end{cases}$$

$X$ 的数学期望 $EX=\frac{a+b}{2}$ . $X^2$ 的数学期望为

$$EX^2=\int_{-\infty}^{+\infty}x^2p(x)\mathrm{d}x=\int_a^b\frac{x^2}{b-a}\mathrm{d}x=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3},$$

得

$$D(X)=EX^2-(EX)^2=\frac{a^2+ab+b^2}{3}-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}.$$

**例 4.12(指数分布的方差)** 设随机变量 $X$ 服从参数为 $\lambda$ 的指数分布 $E(\lambda)$ , 试求 $D(X)$.

**解** 随机变量 $X$ 的密度函数为 $p(x)=\lambda e^{-\lambda x},x>0$ , 且 $X$ 的数学期望 $EX=1/\lambda$. 由于

$$\begin{aligned}EX^2&=\int_{-\infty}^{+\infty}x^2p(x)\mathrm{d}x=\int_0^{+\infty}x^2\lambda e^{-\lambda x}\mathrm{d}x=-x^2e^{-\lambda x}\Big|_0^{+\infty}+2\int_0^{+\infty}xe^{-\lambda x}\\&=\frac{2}{\lambda}\int_0^{+\infty}x\lambda e^{-\lambda x}\mathrm{d}x=\frac{2}{\lambda^2}.\end{aligned}$$

所以 $X$ 的方差 $D(X)=EX^2-(EX)^2=2\lambda^{-2}-\lambda^{-2}=\lambda^{-2}$.

**例 4.13 (正态分布的方差)** 设随机变量 $X$ 服从正态分布 $N(\mu,\sigma^2)$, 试求 $D(X)$.

**解** 随机变量 $X$ 的密度函数为 $p(x) = \dfrac{1}{\sqrt{2\pi}\sigma}e^{(-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$, 数学期望为 $EX = \mu$ , 故其方差

$$D(X) = E(X - EX)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 p(x) \mathrm{d}x = \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x,$$

作变量变换 $t = \frac{x-\mu}{\sigma}$ , 则 $X$ 的方差为

$$D(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}} \mathrm{d}t = \sigma^2 \left[ -\frac{1}{\sqrt{2\pi}} t^2 e^{-\frac{t^2}{2}} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} \mathrm{d}t \right] = \sigma^2.$$

可见 $N(\mu, \sigma^2)$ 中的参数 $\sigma^2$ 等于正态分布的方差. 特别地, 若 $X$ 服从标准正态分布, 则 $X$ 的方差为 1 . 于是我们发现正态分布中的两个参数均有着重要的意义.

### 4.2.2 方差的性质

下面来介绍方差的几个性质. 要注意这些性质与期望的性质之间的差别, 以免引起混淆.

(1) 常数的方差为 0, 即任意常数 $a, D(a) = 0$.

(2) 设 $a, b$ 为任意有限常数, 若 $X$ 的方差存在, 则

$$D(aX + b) = a^2 D(X).$$

**证明** $aX + b$ 的期望为 $E(aX + b) = aEX + b$, 由方差的定义知,

$$D(aX + b) = E[aX + b - (aEX + b)]^2 = E(aX - aEX)^2 = a^2 E(X - EX)^2 = a^2 D(X).$$

(3) 对任意的随机变量 $X$ 和 $Y$, 若 $X$ 和 $Y$ 的方差均存在, 则

$$D(X \pm Y) = D(X) + D(Y) \pm 2E[(X - EX)(Y - EY)].$$

特别地, 当 $X$ 和 $Y$ 独立时,

$$D(X \pm Y) = D(X) + D(Y).$$

**证明** 我们仅证明和的情形, 差的情形可以类似得到.

$$\begin{aligned} D(X + Y) &= E[X + Y - E(X + Y)]^2 = E[(X - EX) + (Y - EY)]^2 \\ &= E[(X - EX)^2 + (Y - EY)^2 + 2(X - EX)(Y - EY)] \\ &= D(X) + D(Y) + 2E[(X - EX)(Y - EY)]. \end{aligned}$$

当 $X$ 和 $Y$ 独立时, 则 $X - EX$ 与 $Y - EY$ 也是独立的. 利用期望的性质 (3),

$$E[(X - EX)(Y - EY)] = E(X - EX)E(Y - EY) = 0,$$

所以

$$D(X+Y)=D(X)+D(Y).$$

上述性质可以推广到 $n$ 个的情形, 任意 $n$ 个随机变量 $X_1,...,X_n$, 则

$$D(X_1+...+X_n)=\sum_{k=1}^{n}D(X_k)+2\sum_{1\leq i<j\leq n}E[(X_i-EX_i)(X_j-EX_j)].$$

进而, 如果 $X_1,...,X_n$ 两两独立, 则

$$D(X_1+...+X_n)=D(X_1)+D(X_2)+...+D(X_n).$$

下面我们看几个例子说明一下方差性质的应用. 首先我们还是运用方差的性质来计算二项分布的方差.

**例 4.20** 试计算二项分布 $B(n,p)$ 的方差.

**解** 设 $X\sim B(n,p)$,, 则 $X$ 可以表示成 $n$ 个独立同分布随机变量的和:

$$X=X_1+X_2+...+X_n,$$

其中,$X_i\sim B(1,p), i=1,2,...,n$ . 于是, 由方差性质 (3), 以及 $X_i$ 的方差为 $pq$ , 得 $X$ 的方差为

$$D(X)=D(X_1)+D(X_2)+...+D(X_n)=np(1-p)=npq.$$

不难看出, 该方法比直接计算要简单得多.

**例 4.21** 设随机变量 $X,Y$ 相互独立, 均服从标准正态分布, 试求 $D(\mid X-Y\mid)$.

**解** 由第三章的内容, 不难看出 $Z=X-Y\sim N(0,2)$. 于是

$$D(\mid X-Y\mid)=D(\mid Z\mid).$$

可得 $D(\mid Z\mid)=E\mid Z\mid^2-(E\mid Z\mid)^2=EZ^2-(E\mid Z\mid)^2$. 再利用定理 4.1, 有

$$\begin{aligned}E\mid Z\mid&=\int_{-\infty}^{+\infty}\mid z\mid\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{z^2}{2\cdot 2}}\mathrm{d}z\\&=\int_{0}^{+\infty}\frac{1}{\sqrt{\pi}}ze^{-\frac{z^2}{4}}\mathrm{d}z\\&=-\int_{0}^{+\infty}\frac{2}{\sqrt{\pi}}\mathrm{d}(e^{-\frac{z^2}{4}})\\&=\frac{1}{\sqrt{\pi}}.\end{aligned}$$

最终得到 $D(\mid Z\mid)=EZ^2-(E\mid Z\mid)^2=2-\left(\frac{2}{\sqrt{\pi}}\right)^2=2-\frac{4}{\pi}$.

### 4.2.3 切比雪夫不等式

下面我们介绍一下切比雪夫不等式, 该不等式在概率论的历史上占据着极为重要的作用. **定理 4.3(切比雪夫不等式)** 设随机变量 $X$ 的期望 $EX$ 和方差 $DX$ 均存在, 则对任意的 $\varepsilon > 0$, 有

$$P(|\ X - EX\ | \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}.$$

**证明** 这里我们仅对随机变量 $X$ 为连续型情况加以证明. 设随机变量 $X$ 的密度函数为 $p(x)$, 则

$$\begin{aligned} P(|\ X - EX\ | \geq \varepsilon) &= \int_{x:|x-EX| \geq \varepsilon} p(x)\mathrm{d}x \leq \int_{x:|x-EX| \geq \varepsilon} \frac{(x-EX)^2}{\varepsilon^2} p(x)\mathrm{d}x \\ &\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x-EX)^2 p(x)\mathrm{d}x = \frac{DX}{\varepsilon^2}. \end{aligned}$$

切比雪夫不等式是一个重要的不等式, 它在大数定律的证明中起着重要的作用. 下面我们举一个例子说明切比雪夫不等式的用处.

由方差的性质 (1) 知, 常数的方差为零; 反之, 若随机变量 $X$ 的方差为零, 该随机变量是否为常数? 直观上看结果是显然的, 但如何证明呢? 事实上我们可以运用切比雪夫不等式得到严格的证明. 假设 $D(X) = 0$, 对任意的 $\varepsilon > 0$,

$$0 \leq P(|\ X - EX\ | \geq \varepsilon) \leq \frac{\mathrm{D}X}{\varepsilon^2} = 0,$$

即 $P(|\ X - EX\ | \geq \varepsilon) = 0$. 从而

$$\begin{aligned} P(X \neq EX) &= P\left(\bigcap_{i=1}^{\infty} \left\{|\ X - EX\ | \geq \frac{1}{n}\right\}\right) \\ &= \lim_{n\to\infty} P\left(|\ X - EX\ | \geq \frac{1}{n}\right) = 0. \end{aligned}$$

即 $P(X = EX) = 1$. 也就是说, 随机变量 $X$ 几乎处处为一常数. 这样我们就给出了证明.

## 4.3 协方差与相关系数

前面两节我们学习了随机变量的数字特征. 这一节将介绍随机向量的数字特征. 我们知道, 在第三章表示随机向量之间的关系有条件分布, 独立性等. 那么如何用数字来描述两个随机变量之间的关系呢? 这就需要相关系数的概念. 首先我们介绍一下协方差的概念.

### 4.3.1 协方差

**定义 4.4** 对随机变量 $X$ 和 $Y$, 若 $EX, EY$ 和 $E(XY)$ 均存在, 则称 $E[(X-EX)(Y-EY)]$ 为 $X$ 和 $Y$ 的协方差, 记为 $\mathrm{cov}(X,Y)$. 即

$$\mathrm{cov}(X,Y) = E[(X-EX)(Y-EY)].$$

协方差的概念可以看成方差概念的自然推广. 事实上,$\mathrm{D}X=\mathrm{cov}(X,X)$. 也就是说, 随机变量的方差相当于自己和自己的协方差. 由方差的性质 (3), 不难看出:

$$D(X\pm Y)=D(X)+D(Y)\pm 2\mathrm{cov}(X,Y).$$

类似地, 对任意 $n$ 个随机变量 $X_1,...,X_n$, 有

$$D(X_1+...+X_n)=\sum_{k=1}^{n}D(X_k)+2\sum_{1\leq i<j\leq n}\mathrm{cov}(X_{ij},X_j).$$

此外, 将协方差的定义展开得

$$\mathrm{cov}(X,Y)=E[(X-EX)(Y-EY)]=EXY-E[XEY]-E[YEX]+EX\cdot EY.$$

注意到后面两项可以消去, 于是得到下列协方差的计算公式:

$$\mathrm{cov}(X,Y)=E(XY)-EX\cdot EY.$$

类似地, 协方差有如下的性质:
(1) 对称性,$\mathrm{cov}(X,Y)$ =$\mathrm{cov}(Y,X)$.
(2) 对常数 $a,b,c,d\in\mathbb{R}$ ,$\mathrm{cov}(aX+c,bY+d)=ab\mathrm{cov}(X,Y)$.
(3)$\mathrm{cov}(X_1+X_2,Y)$ =$\mathrm{cov}(X_1,Y)$+$\mathrm{cov}(X_2,Y)$.
(4) 若 $X$ 和 $Y$ 独立, 则 $\mathrm{cov}(X,Y)=0$.
证明 我们仅对性质 (4) 加以证明, 由

$$\mathrm{cov}(X,Y)=E(XY)-EX\cdot EY.$$

利用期望的性质 (3), 得 $\mathrm{cov}(X,Y)=0$.
**例 4.14** 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 求 $\mathrm{cov}(X,Y)$.
**解** 由第三章, 二维正态分布的密度函数为

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

且 $EX=\mu_1,EY=\mu_2$, 于是

$$\mathrm{cov}(X,Y)=E[(X-EX)(Y-EY)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)p(x,y)\mathrm{d}x\mathrm{d}y.$$

对积分进行变量变换,$u=\dfrac{x-\mu_1}{\sigma_1}, v=\dfrac{y-\mu_2}{\sigma_2}$, 得

$$
\begin{aligned}
\operatorname{cov}(X,Y) &= \frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} uv\exp\left\{-\frac{1}{2(1-\rho^2)}[(u-\rho v)^2+(1-\rho^2)v^2]\right\}\mathrm{d}u\mathrm{d}v \\
&= \frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} ve^{-\frac{v^2}{2}}\mathrm{d}v\int_{-\infty}^{+\infty} u\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left[\frac{(u-\rho v)^2}{2(1-\rho^2)}\right]\mathrm{d}u
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} v^2e^{-\frac{v^2}{2}}\mathrm{d}v \\
&= \sigma_1\sigma_2\rho.
\end{aligned}
$$

最后我们介绍重要的柯西-施瓦茨 (Cauchy-Schwarz) 不等式. 该不等式有多种形式, 如积分形式, 数列形式等. 我们这里介绍的是概率论中的形式, 事实上, 它们的证明与结论都是类似的.

**定理 4.4(柯西-施瓦茨不等式)** 设随机变量 $X$ 和 $Y$ 的方差均存在, 则

$$[\operatorname{cov}(X,Y)]^2 \leq D(X)D(Y),$$

其中等号成立的充要条件是存在不全为 0 的常数 $a$ 和 $b$, 使得

$$P(Y=aX+b)=1.$$

**证明** 对任意的 $t\in\mathbb{R}$, 定义函数

$$u(t)=E[t(X-EX)-(Y-EY)]^2=t^2D(X)-2t\operatorname{cov}(X,Y)+D(Y),$$

显然 $u(t)\geq 0$. 由判别式知 $[2\operatorname{cov}(X,Y)]^2-4D(X)D(Y)\leq 0$, 即

$$[\operatorname{cov}(X,Y)]^2\leq D(X)D(Y).$$

等式成立的充要条件是 $u(t)=0$ 有一个二重实根, 设其为 $t_0$. 于是

$$u(t_0)=E[t_0(X-EX)-(Y-EY)]^2=D(t_0(X-EX)-(Y-EY)]=0,$$

由上一节切比雪夫不等式后的论述知,

$$P(t_0(X-EX)-(Y-EY)=0)=1.$$

等式的必要性得证.

等式部分的充分性是显然的, 若存在不全为 0 的常数 $a,b$, 使得 $P(Y=aX+b)=1$. 于是

$$\begin{aligned}\operatorname{cov}(X,Y)&=E(XY)-EX\cdot EY=E[X(aX+b)]-EX\cdot E(aX+b)\\&=aE(X^2)-a(EX)^2=\sqrt{X\cdot a^2DX}=\sqrt{DX\cdot DY}.\end{aligned}$$

等式的充分性得证.

### 4.3.2 相关系数

在讲述相关系数之前, 我们先介绍一下标准化的概念. 在已知随机变量 $X$ 的均值 $EX$ 和方差 $DX$ 时, 我们可以进行如下变换:

$$X^*=\frac{X-EX}{\sqrt{D(X)}}.$$

不难看出

$$EX^*=E\frac{X-EX}{\sqrt{D(X)}}=0,\qquad DX^*=D\left(\frac{X-EX}{\sqrt{D(X)}}\right)=\frac{E(X-EX)^2}{D(X)}=1.$$

则称 $X^*$ 为 $X$ 的**标准化随机变量.** 例如, 若 $X\sim N(\mu,\sigma^2)$, 则其标准化随机变量 $\dfrac{X-\mu}{\sigma}\sim N(0,1)$, 也就是说, 正态分布标准化后就服从标准正态分布了.

下面介绍相关系数的概念.

**定义 4.5** 设随机变量 $X$ 和 $Y$ 的方差均存在, 且 $D(X)>0$,$D(Y)>0$, 则称

$$\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)D(Y)}}$$

为 **$X$ 和 $Y$ 的相关系数**, 记为 $\rho_{XY}$ 或者 $\operatorname{Corr}(X,Y)$. 若 $\rho_{XY}>0$, 则称 $X$ 和 $Y$ 正相关; 若 $\rho_{XY}<0$,, 则称 $X$ 和 $Y$ 负相关.

考虑 $X$ 和 $Y$ 的标准化随机变量 $X^*=\dfrac{X-EX}{\sqrt(D(X)}$ 和 $Y^*=\dfrac{Y-EY}{\sqrt(D(Y)}$, 则

$$\operatorname{cov}(X^*,Y^*)=\operatorname{cov}\left(\frac{X-EX}{\sqrt{D(X)}},\frac{Y-EY}{\sqrt{D(Y)}}\right)=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)D(Y)}}=\rho_{XY}.$$

可见相关系数是随机变量标准化后的协方差. 由柯西-施瓦茨不等式, 可知相关系数 $\rho_{XY}$ 满足下列性质:

(1)$|\rho_{XY}|\le 1$;

(2)$|\rho_{XY}|=1$ 的充分必要条件是存在不同为 0 的常数 $a,b$, 使得 $P(Y=aX+b)=1$. 即 $X$ 和 $Y$ 以概率 1 具有线性关系.

上面的性质表明, 相关系数 $\rho_{XY}$ 刻画了 $X$ 和 $Y$ 间的线性相关特征.$|\rho_{XY}|$ 越大, 表明 $X$ 和 $Y$ 之间线性关系越密切, 当 $|\rho_{XY}|=1$ 时表明 $X$ 和 $Y$ 以概率 1 线性相关. 反之, 若 $|\rho_{XY}|$ 越小, 表明 $X$ 和 $Y$ 的线性关系越弱. 特别地, 对于 $\rho_{XY}=0$ 情形, 我们给出如下的定义.

**定义 4.6** 若随机变量 $X$ 和 $Y$ 的相关系数 $\rho_{XY}=0$, 则称 $X$ 和 $Y$ **线性无关**或**不相关**.

(3) 若随机变量 $X$ 和 $Y$ 独立, 则 $X$ 和 $Y$ 不相关. 但反之未必成立.

**证明** 若随机变量 $X$ 和 $Y$ 独立, 则由期望的性质 (3) 得 $E(XY) = EX \cdot EY$, 于是 $\operatorname{cov}(X,Y) = E(XY) - EX \cdot EY = 0$,, 所以 $\rho_{XY} = 0$.

独立和不相关并非等价, 不相关比独立的条件要弱一些. 下面我们再举一个例子来说明这个问题.

**例 4.15** 设 $X \sim N(0,1)$,$Y = X^2$, 试求 $\rho_{XY}$.

**解** 因为 $EX = 0$, $E(XY) = \int_{-\infty}^{+\infty} x^3\varphi(x)\mathrm{d}x = 0$, 所以

$$\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y) = 0,$$

进而 $\rho_{XY} = 0$.

在该例中, 虽然 $X$ 和 $Y$ 不相关, 但事实上它们之间有着严格的非线性函数关系, 所以不独立. 于是我们就明白相关系数 $\rho_{XY}$ 刻画了 $X$ 和 $Y$ 之间的线性相关特征这句话的含义了. 即使相关系数 $\rho_{XY}$ 为零, 也只能说明 $X$ 和 $Y$ 没有线性关系, 也许 $X$ 和 $Y$ 有着很强的非线性关系呢！而独立性则说明随机变量 $X$ 和 $Y$ 之间没有任何关系.

下面的性质给出了 $X$ 和 $Y$ 不相关的几个等价条件:

(4) 下面四个命题相互等价:

(a)$\rho_{XY} = 0$;

(b)$\operatorname{cov}(X,Y) = 0$;

(c)$E(XY) = EX \cdot EY$;

(d)$D(X \pm Y) = D(X) + D(Y)$.

一般地, 独立性与不相关是不等价的, 但在正态分布的情况下, 独立性与不相关却是等价的.

**例 4.16** 设 $(X,Y)$ 服从二维正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则它们的独立性和不相关性等价.

**证明** 由 $\operatorname{cov}(X,Y) = \sigma_1\sigma_2\rho$, 所以 $X$ 和 $Y$ 的相关系数为

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)D(Y)}} = \frac{\sigma_1\sigma_2\rho}{\sigma_1\sigma_2} = \rho.$$

于是我们发现二维正态分布中参数 $\rho$ 的含义了, 它就是随机变量 $X$ 和 $Y$ 的相关系数. 由第三章的定理 3.2 知,$X$ 和 $Y$ 独立等价于 $\rho = 0$. 于是 $X$ 和 $Y$ 的独立性与不相关性等价.

## 4.4 矩与协方差阵

### 4.4.1 矩

作为期望和方差的推广, 在本节我们将介绍矩的概念. 对于随机变量 $X$ 和正整数 $k$, 我们定义 $X^*$ 的数学期望 $EX^*$ 为随机变量 $X$ 的 $k$ 阶矩. 若 $k = 1$, 则 $EX$ 就是 $X$ 的期望. 因此, 矩可以看作期望的推广; 同样, 可以定义中心化后随机变量的期望 $E(X - EX)^k$. 若 $k = 2$, 则 $E(X - EX)^2$ 就是 $X$ 的方差. 下面我们给出具体的定义.

**定义 4.7** 设 $X$ 是一个随机变量, 对正整数 $k$, 若 $E(\mid X \mid^k) < \infty$ , 则称 $EX^k$ 为 $X$ 的 $k$ 阶**原点矩**或 $k$ **阶矩**; 若 $E(\mid X - EX \mid^k) < \infty$ , 则称 $E(X - EX)^k$ 为 X 的 $k$ 阶**中心矩**.

同样, 随机变量的 $k$ 阶矩也不一定存在. 若 $X$ 的 $k$ 阶矩或 $k$ 阶中心矩存在, 则可以运用随机变量函数的期望计算公式 (即定理 4.1) 计算 $k$ 阶矩及 $k$ 阶中心矩.

**例 4.17** 设 $X \sim N(\mu, \sigma^2)$, 求 $X$ 的 $k$ 阶中心矩, 其中 $k$ 为任意正整数.

**解** 随机变量 $X$ 的密度函数为 $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{(-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$, 数学期望为 $EX = \mu$. 故其 $k$ 阶中心矩为

$$E(X-EX)^k = \int_{\infty}^{+\infty} (x-\mu)^k \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \int_{\infty}^{+\infty} \sigma^k t^k \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

其中第二个等号还是利用变量变换 $t = \dfrac{x-\mu}{\sigma}$.

当 $k$ 为奇数时, 被积函数是一个可积的奇函数, 所以 $E(X-EX)^k = 0$.

当 $k$ 为偶数时, 有

$$\begin{aligned}
E(X-EX)^k &= \sigma^k \int_{-\infty}^{+\infty} t^k \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= -\sigma^k \int_{-\infty}^{+\infty} t^{k-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= -\frac{\sigma^k}{\sqrt{2\pi}} \left[ t^{k-1} e^{-\frac{t^2}{2}} \right]_{-\infty}^{+\infty} + \sigma^k (k-1) \int_{-\infty}^{+\infty} t^{k-2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= \sigma^k (k-1) \int_{-\infty}^{+\infty} t^{k-2} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= \cdots = \sigma^k (k-1)(k-3) \cdots 3 \cdot 1 \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \\
&= \sigma^k (k-1)(k-3) \cdots 3 \cdot 1
\end{aligned}$$

综上所述, 得

$$E(X-EX)^k = \begin{cases} 0, k\text{为奇数} \\ \sigma^k (k-1)!!, k\text{为偶数}. \end{cases}$$

我们将随机变量标准化后的三阶矩称为偏度, 同时将标准化随机变量的四阶矩称为峰度. 由例 4.17 不难看出, 正态分布的偏度为 0, 峰度为 3. 峰度可以用来衡量分布形状的陡峭程度. 在实际中, 高于四阶的矩是较少用到的.

### 4.4.2 * 协方差阵

对于 $n$ 维随机向量 $X = (X_1, X_2, ..., X_n)^T$, 也可以定义其数学期望及协方差阵.

**定义 4.8** 对随机向量 $X = (X_1, X_2, ..., X_n)^T$, 称

$$EX = (EX_1, EX_2, ..., EX_n)^T$$

为 $X$ 的**数学期望**.

记 $c_{ij} = \mathrm{cov}(X_i, X_j), i, j = 1, 2, ..., n$. 称矩阵

$$\Sigma = (c_{ij})_{n \times n} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

为 $X$ 的**协方差阵**.

显然, 协方差阵是一个对称阵. 例如, 我们考虑二维正态分布 $(X_1, X_2)$ 时, 其协方差阵为

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

此时二维正态分布的联合密度函数可以写成如下形式:

$$p(x) = \frac{1}{(2\pi)^{\frac{2}{2}}(\det\Sigma)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right],$$

其中 $x = (x_1, x_2)^T$,$\mu = (\mu_1, \mu_2)^T = (EX_1, EX_2)^T$. 该公式展开后与第三章公式是完全相同的. 且容易记忆, 因为它与一维正态分布的公式是类似的.

# 5 极限理论

极限理论是概率论的重要内容, 在概率统计的理论研究与实际应用中具有非常重要的意义. 极限理论主要包含大数定律与中心极限定理, 它反映了随机变量序列的频率稳定性与分布稳定性问题.

第一章中我们提出了频率的稳定性问题, 即随着试验次数 $n$ 的增加, 事件的频率将在某个数值 $p$ 附近稳定地摆动. 而稳定性究竟怎样描述, 如何用数学语言把它表达出来, 大数定律从理论上阐述了这一问题.

大量随机因素的叠加会产生正态分布的特征, 中心极限定理从理论上揭示了这一现象.

## 5.1 大数定律

大数定律是研究随机现象统计规律性的理论, 大数定律揭示了随机变量平均值的收敛规律, 我们先介绍随机变量收敛的概念.

**定义 5.1** 设 $Y_1, Y_2, \cdots, Y_n, \cdots$ 为随机变量序列, 若存在常数 $a$, 对任意给定的 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P(\mid Y_n - a \mid \geq \varepsilon) = 0,$$

或

$$\lim_{n\to\infty} P(\mid Y_n - a \mid < \varepsilon) = 1,$$

则称随机变量序列 $Y_1, Y_2, \cdots, Y_n \cdots$ 依概率收敛于 $a$, 记为 $Y_n \xrightarrow{P} Pa$.

随机变量序列的收敛性是指 $n \to \infty$ 时, 随机变量 $Y_n$ 与常数 $a$ 的偏差小于任意正数 $\varepsilon$ 的概率将趋向于 1 , 这与数学分析中的极限定义有明显不同.

**定义 5.2** 设 $X_1, X_2, \cdots, X_n, \cdots$ 为随机变量序列, 若

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| \geq \varepsilon\right) = 0.$$

或

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \varepsilon\right) = 1.$$

则称 $\{X_n\}$ 服从大数定律.

不同的大数定律主要是条件的不同. 以下我们介绍三个常用的大数定律: 切比雪夫大数定律, 独立同分布大数定律与伯努利大数定律.

**定理 5.1(切比雪夫大数定律)** 设 $X_1, X_2, \cdots$ 为两两互不相关的随机变量序列, 其方差一致有界, 即存在常数 $C$, 使得 $DX_k < C$, 对一切 $k = 1, 2, \cdots$ 成立, 则 $\{X_n\}$ 服从大数定律, 即

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| \geq \varepsilon\right) = 0.$$

或

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} EX_k.$$

**证明** 先证 $\dfrac{1}{n^2}D\left(\sum\limits_{k=1}^{n}X_k\right)\to 0$. 由于 $\{X_k\}$ 两两互不相关, 即 $\operatorname{cov}(X_i,X_j)=0$, 所以

$$\begin{aligned}\frac{1}{n^2}D\left(\sum_{k=1}^{n}X_k\right)&=\frac{1}{n^2}\left[\sum_{k=1}^{n}DX_k+2\sum_{1\le i<j\le n}\operatorname{cov}(X_i,X_j)\right]\\&=\frac{1}{n^2}\sum_{k=1}^{n}DX_k<\frac{1}{n^2}\sum_{k=1}^{n}C=\frac{C}{n}\to 0.\end{aligned}$$

根据切比雪夫不等式

$$\begin{aligned}0&\le P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}EX_k\right|\ge\varepsilon\right)=P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}E\left(\sum_{k=1}^{n}X_k\right)\right|\ge\varepsilon\right)\\&\le D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)/\varepsilon^2=\frac{1}{n^2}D\left(\sum_{k=1}^{n}X_k\right)/\varepsilon^2\to 0.\end{aligned}$$

**定理 5.2(独立同分布大数定律)** 设 $X_1,X_2,\cdots$ 为相互独立且分布相同的随机变量序列, 数学期望存在,$EX_n=\mu$, 则 $\{X_n\}$ 服从大数定律, 即

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

**证明** 只在方差存在的情况下给出证明.

因为 $\{X_n\}$ 同分布, 设 $DX_n=\sigma^2<\infty$, 从而方差一致有界, 由于 $X_1,X_2,\cdots$ 相互独立, 必有两两互不相关, 而 $\dfrac{1}{n}\sum\limits_{k=1}^{n}EX_k=\mu$, 根据切比雪夫大数定律

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

因此, 大数定律成立. 可以表示为 $\dfrac{1}{n}\sum\limits_{k=1}^{n}X_k\xrightarrow{P}\mu$.

大数定律给出了频率稳定性的严格数学定义, 即大量独立随机观测的平均值依概率收敛于分布的期望值. 因此, 对于一些具有随机性的测量结果, 以多次测量的平均值作为测量值会更加准确. 这也为估计任意分布的随机变量的数学期望提供了可行方法.

**例 5.1** 以上述思想方法, 计算定积分 $\displaystyle\int_0^1 f(x)\mathrm{d}x$.

**解** 设 $X$ 服从 $[0,1]$ 上的均匀分布, 密度 $p(x)=1,0\le x\le 1$, 则

$$E[f(X)]=\int_{-\infty}^{+\infty}f(x)p(x)\mathrm{d}x=\int_0^1 f(x)\mathrm{d}x.$$

根据大数定律,$n$ 很大时,$f(X)$ 的 $n$ 个观测值的平均值接近于期望值 $E[f(X)]$, 即 $E[f(X)]\approx\dfrac{1}{n}\sum\limits_{k=1}^{n}f(x_k)$. 具体求法: 在计算机上用随机数发生器产生 $n$ 个 $[0,1]$ 区间上均匀分布的随机数 $x_1,x_2,\cdots,x_n$, 便有

$\int_0^1 f(x)dx \approx \frac{1}{n}\sum_{k=1}^{n} f(x_k)$. 算例,$\int_0^1 e^{x^2}\mathrm{d}x$ 准确值为 1.4626517..., 取 $n = 10^4$, 近似值为 1.46285; 取 $n = 10^5$, 近似值为 1.46268.

**例 5.2** 设 $X_1, X_2, \cdots, X_n$ 独立同分布, 均服从参数为 $\lambda$ 的指数分布, 当 $n$ 趋于无穷大时, 问 $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k^2$ 依概率收敛于何值?

**解** 设 $X \sim E(\lambda), X_1, X_2, \cdots, X_n$ 相互独立与 $X$ 同分布, 根据独立性的性质, 可得 $X_1^2, X_2^2, \cdots, X_n^2$ 相互独立且分布相同. 由独立同分布大数定律,

$$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k^2 \xrightarrow{P} EX^2.$$

$X \sim E(\lambda)$, 因此 $EX = \frac{1}{\lambda}, DX = \frac{1}{\lambda^2}$,

$$EX^2 = DX + (EX)^2 = \frac{1}{\lambda^2} + \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2},$$

$$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k^2 \xrightarrow{P} \frac{2}{\lambda^2}.$$

**定理 5.3(伯努利大数定律)** 设 $n_A$ 是 $n$ 重伯努利试验中 $A$ 发生的次数,$A$ 发生的概率为 $p$, 则对任意 $\varepsilon > 0$, 有

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

**证明** 令

$$X_k = \begin{cases} 1, & \text{第 k 次实验中 A 发生} \\ 0, & \text{第 k 次实验中 A 不发生} \end{cases}, k = 1, 2, \cdots, n.$$

则 $X_1, X_2, \cdots, X_n$ 相互独立, 均服从 $0-1$ 分布,$EX_k = p, \frac{1}{n}\sum_{k=1}^{n} EX_k = p$.

显然,$n_A = X_1 + X_2 + \cdots + X_n, \frac{n_A}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k$, 由独立同分布大数定律

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right) = \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| \geq \varepsilon\right) = 0.$$

或

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1.$$

因此伯努利大数定律成立, 可以表示为 $\frac{n_A}{n} \xrightarrow{P} p$.

伯努利大数定律表明, 当试验次数 $n$ 不断增加时, 事件 $A$ 发生的频率与 $A$ 发生的概率 $p$ 的偏差小于任意正数 $\varepsilon$ 的概率趋向于 1, 这从理论上说明了频率的稳定性. 从实用的角度看, 当试验次数充分大时, 可以用事件发生的频率代替事件发生的概率.

## 5.2 中心极限定理

有一类随机变量, 由大量独立随机因素叠加而成, 其中每一个因素的影响是微小的, 不占主导地位的. 那么, 这样的随机变量其分布会呈现出正态分布的特征.

**定义 5.3(标准化随机变量)** 若随机变量 $X$ 满足

$$EX=0, DX=1,$$

则称 $X$ 为标准化随机变量.

对于任意随机变量 $X$, 令 $Y=\dfrac{X-E(X)}{\sqrt{D(X)}}$, 则 $Y$ 为标准化随机变量.

**定理 5.4** 独立同分布中心极限定理 (林德贝格-列维) 随机变量序列 $X_1, X_2, \cdots, X_n$ 独立同分布, 数学期望与方差均存在,$EX_k=\mu, DX_k=\sigma^2>0$, 则对任意 $x$ 有

$$\lim_{n\to\infty} P\left(\frac{\sum\limits_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t = \Phi(x).$$

证明略.

设上述随机变量序列和 $\sum\limits_{k=1}^{n} X_k$ 的标准化随机变量为 $Y_n$, 则

$$Y_n = \frac{\sum\limits_{k=1}^{n} X_k - E\left(\sum\limits_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n} X_k\right)}} = \frac{\sum\limits_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}.$$

定理 5.4 说明,$Y_n$ 的分布函数

$$F_n(x) = P(Y_n \leq x) = P\left(\frac{\sum\limits_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leq x\right).$$

的极限形式是 $\Phi(x)$, 也就是说随机变量之和的标准化随机变量 $Y_n$ 的极限分布是标准正态分布. 在实际问题中, 一般当 $n$ 较大时, 可以认为独立同分布随机变量的和近似服从正态分布

$$\frac{\sum\limits_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \sim N(0,1),$$

或

$$\sum_{k=1}^{n} X_k \sim N(n\mu, n\sigma^2),$$

即

$$\sum_{k=1}^{n} X_k \sim N\left[E\left(\sum_{k=1}^{n} Xk\right), D\left(\sum_{k=1}^{n} Xk\right)\right].$$

一般而言,$n$ 个随机变量之和的概率分布, 即使已知每个随机变量的分布也是很难求出的, 而利用中心极限定理, 当 $n$ 很大时, 可以用正态分布来近似描述 $\sum\limits_{k=1}^{n} X_k$ 的分布, 这为 $k=1$ 解决很多实际问题提供了方便.

**例 5.3** 一信号接收器同时收到 20 个信号电压, 设它们互相独立均服从均匀分布 $U(0,10)$, 求电压之和大于 105 的概率.

**解** 记信号电压为 $V_k$, 则 $EV_k = 5, DV_k = 10^2/12, k = 1, 2, \cdots, 20$.

设 $V = \sum\limits_{k=1}^{n} V_k$, 则 $EV = \sum\limits_{k=1}^{n} EV_k = 100, DV = \sum\limits_{k=1}^{n} DV_k = \dfrac{500}{3}$, 由中心极限定理,$V \sim \left(100, \frac{500}{3}\right)$ ,$\frac{V-100}{\sqrt{500/3}} \sim N(0,1)$.

$$P(V > 105) = P\left(\frac{V-100}{\sqrt{500/3}} > \frac{105-100}{\sqrt{500/3}}\right) \approx 1 - \Phi(0.39) = 0.3483.$$

**例 5.4** 某元件的使用寿命 (单位: 小时) 服从指数分布, 已知其平均寿命为 20 小时. 某仪器不间断地使用该元件, 一只损坏后自动更换另一只. 问应至少准备多少只, 才能以 95% 以上的概率保证可连续使用 2000 小时.

**解** 设至少准备 $n$ 只, 第 $k$ 只寿命记为 $X_k, X_k \sim E(\lambda)$, 平均寿命 $EX_k = 1/\lambda = 20$, 所以, $DX_k = 1/\lambda^2 = 400$, 则 $\sum\limits_{k=1}^{n} X_k = 20n, D\sum\limits_{k=1}^{n} X_k = 400n$, 由中心极限定理,

$$\sum_{k=1}^{n} X_k \sim N(20n, 400n),$$

从而

$$\begin{aligned} P\left(\sum_{k=1}^{n} X_k \geq 2000\right) &= P\left(\frac{\sum\limits_{k=1}^{n} X_k - 20n}{\sqrt{400n}} \geq \frac{2000-20n}{\sqrt{400n}}\right) \\ &= 1 - \Phi\left(\frac{2000-20n}{\sqrt{400n}}\right) = \Phi\left(-\frac{2000-20n}{\sqrt{400n}}\right) > 0.95, \end{aligned}$$

查表,$\Phi(1.64) = 0.95, -\dfrac{2000-20n}{\sqrt{400n}} > 1.64$, 即 $20n - 32.8\sqrt{n} - 2000 > 0$, 解之,$n > 117.8$ 或 $n < 85$(舍), 所以应至少准备 118 只.

**例 5.5** 由大数定律我们知道, 独立同分布随机变量 $\{X_k\}$ 的平均值依概率收敛于其均值 $\mu$, 设 $n$ 很

大, 试用中心极限定理求 $\frac{1}{n}\sum_{k=1}^{n}X_k$ 与 $\mu$ 的偏差小于 $\varepsilon$ 的概率.

**解** 设 $EX_k=\mu, DX_k=\sigma^2$, 由中心极限定理

$$\sum_{k=1}^{n}X_k \sim N(n\mu, n\sigma^2),$$

从而

$$N(0,1)\frac{\frac{1}{n}\sum_{k=1}^{n}X_k-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

所以有

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right)=P\left(\left|\frac{\frac{1}{n}\sum_{k=1}^{n}X_k-\mu}{\sigma/\sqrt{n}}\right|<\frac{\varepsilon}{\sigma/\sqrt{n}}\right)=2\Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right)-1.$$

**定理 5.5(拉普拉斯中心极限定理)** 设 $\mu_n$ 是 $n$ 重伯努利试验中 $A$ 发生的次数, 每次试验中 $A$ 发生的概率为 $p, q=1-p$, 则对任意 $x$ 有

$$\lim_{n\to\infty}P\left(\frac{\mu_n-np}{\sqrt{npq}}\leq x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}\mathrm{d}t=\Phi(x).$$

**证明** 令

$$X_k=\begin{cases}1, & \text{第 k 次实验中 A 发生}\\ 0, & \text{第 k 次实验中 A 不发生}\end{cases}, k=1,2,\cdots,n.$$

则 $X_1, X_2, ..., X_n$ 相互独立, 均服从 0-1 分布,$\mu_n=\sum_{k=1}^{n}X_k$, 且 $EX_k=p, DX_k=\sigma^2=pq$, 代入独立同分布中心极限定理, 可见定理 5.4 成立.

显然, 拉普拉斯中心极限定理中的 $\mu_n$ 服从二项分布 $B(n,p)$, 定理表明,$\mu_n$ 的标准化随机变量 $\frac{\mu_n-np}{\sqrt{npq}}$, 其极限分布为标准正态分布. 当 $n$ 充分大时,$\frac{\mu_n-np}{\sqrt{npq}}$ 近似服从标准正态分布. 等价地,$n$ 充分大时 $\mu_n \sim N(np, npq)$.

**这说明, 服从二项分布的随机变量,$n$ 充分大时近似服从正态分布.**

这一性质, 也为二项分布的计算带来方便. 例如, 按二项分布计算

$$P(a\leq\mu_n\leq b)=\sum_{a\leq k\leq b}C_n^k p^k q^{n-k},$$

当求和项数很多时, 计算复杂. 而用中心极限定理, 计算简单.

$$
\begin{aligned}
P(a \leq \mu_n \leq b) &= P(\frac{a-np}{\sqrt{npq}} \leq \frac{\mu_n - np}{\sqrt{npq}} \leq \frac{b-np}{\sqrt{npq}}) \\
&= \Phi(\frac{b-np}{\sqrt{npq}}) - \Phi(\frac{a-np}{\sqrt{npq}}).
\end{aligned}
$$

# 6　统计量与抽样分布

前面我们介绍了概率论的基本内容. 从本章起, 我们将进入本课程的另外一个部分-数理统计学. 数理统计学也是一门研究随机现象的学科, 它以概率论为其理论基础, 但它的研究内容及方式与概率论又有所不同. 通过概率论的学习, 可以看到概率论有着严谨的理论体系, 从公理到定理, 以演绎的方式展开; 而数理统计则比较注重应用, 它的理论方法往往更具有归纳的特点.

粗略地讲, 数理统计是一门研究带有随机因素数据的学科. 它的内容大致分为两类: 一是研究如何有效地收集随机数据, 如抽样调查; 二是研究如何有效地分析已获得的随机数据, 这类问题包含众多的分支, 如参数估计, 假设检验等. 本书主要就是介绍这方面的基本内容.

数理统计的应用十分广泛. 事实上, 它几乎渗透到人类活动的一切领域. 例如, 农业, 生物领域的“生物统计”; 经济, 商业领域的“计量统计”; 金融领域的“保险统计”; 气象领域的“气象统计”; 医学领域的“医学统计”; 教育和心理学领域的“教育统计”等. 这些领域的统计方法的基础就是数理统计.

## 6.1　总体与样本

### 6.1.1　总体

在数理统计中, 我们将所研究的对象的全体称为**总体**. 而将总体中每个成员称为**个体**. 例如, 我们研究一家工厂的某种产品的废品率. 这种产品的全体就是我们的总体, 而每件产品则是个体. 又如, 我们要研究某市小学生的身高与体重. 那么该市全体小学生的身高与体重就是我们的总体, 而每个小学生的身高与体重就是个体. 实际上, 我们真正关心的并不是总体或个体本身, 而是它们的某一项或几项数量指标 $X$.

**例 6.1** 现用一把尺子测量一个物体的长度. 假定进行了 $n$ 次测量, 得到一组数值

$$x_1, x_2, \cdots, x_n.$$

在这个问题中, 我们可以将总体理解为一切可能的测量值的全体. 由于各种随机的因素, 在不同的测量中得到的测量值是不同的. 因此, 我们通常将总体的数量指标 $X$ 看成随机变量, 它的分布称为总体的分布. 在上例中, 假定物体的真正长度为 $\mu$(未知). 如果测量过程没有系统误差, 可以认为总体 $X$ 服从 $N(\mu, \sigma^2)$. 其中 $\sigma^2$ 反映了测量的精度, 在某些情况下也可以设 $\sigma^2$ 为已知.

如果一个总体所包含的个体数量是有限的, 则称之为有限总体. 如果总体所包含的个体数量是无限的, 则称之为无限总体. 在实际问题中, 即使是有限总体, 其数量也是比较大的, 我们往往将它近似地看作无限总体. 例如, 研究某地区农民的收入. 尽管该地区农民的数量是有限的, 我们仍然将其看成无限总体, 甚至还可以将总体分布取成连续型分布.

### 6.1.2　样本

为了对总体 $X$ 进行研究, 通常要从总体中随机地抽取一些个体, 这些个体就称为**样本**. 抽得样本的过程称为**抽样**. 样本中个体的数量称为**样本容量.**

设对总体进行了 $n$ 次观测, 得到一组数据 $(x_1, x_2, \cdots, x_n)$. 我们称这组数据为**样本观察值**或**样本值**. 统计学的工作就是利用样本观察值对总体分布中的未知成分进行推断. 如例 6.1 中, 我们需要研究如何利用样本观察值对物体的长度 $\mu$ 进行合理的估计. 值得注意的是, 样本具有二重性. 就一次具体的

观察而言,$(x_1,x_2,\cdots,x_n)$ 是完全确定的数, 也就是说, 样本具有数的属性; 但另一方面, 由于各种随机因素的影响, 在不同的观察中样本的取值会发生变化. 因此, 又应该将它们看成随机变量, 即样本同时具有随机变量的属性. 总而言之, 在具体计算中, 我们通常将样本看成一组数, 通常用小写字母表示, 记为 $(x_1,x_2,\cdots,x_n)$. 而在考虑一般问题时, 我们谈到样本, 往往将其看做一组随机变量, 通常用大写字母表示, 记为 $(X_1,X_2,\cdots,X_n)$.

既然样本 $(X_1,X_2,\cdots,X_n)$ 被看做随机变量, 自然需要研究其分布. 为了使样本能很好地反映总体的特征, 我们对随机抽样提出如下要求:

(1) 代表性: 样本能够代表总体. 也就是说, 样本的每个分量 $X_i$ 与总体 $X$ 具有相同的分布.

(2) 独立性:$X_1,X_2,\cdots,X_n$ 为相互独立的随机变量.

满足上述两点性质的样本称为

简单随机样本, 也简称为样本. 今后如不作特别声明, 本书提及的样本均指简单随机样本.

设总体 $X$ 的分布函数为 $F(x)$, 则样本 $(X_1,X_2,\cdots,X_n)$ 的分布函数为

$$F(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F(x_i).$$

若总体是连续性随机变量, 其概率密度函数为 $p(x)$, 则样本 $(X_1,X_2,\cdots,X_n)$ 的概率密度函数为

$$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}p(x_i).$$

在数理统计中, 总体或者说总体分布是我们研究的对象, 而样本是从总体中随机抽取的一部分个体, 这些样本中包含着总体的信息, 通过对这些样本进行具体的研究, 我们可以得到对总体的解释. 因此, 如何利用样本去合理推断总体, 这是数理统计这门学科需要研究的问题. 这种由已知推断未知, 由具体推断抽象的思想, 对我们今后的研究是大有启发的.

## 6.2 统计量与抽样分布

在获得样本之后, 我们就要对总体的未知成分进行推断. 这需要对样本进行加工, 整理, 从中提取有用的信息. 为此要针对不同的问题构造不同的样本函数, 这种函数称为统计量.

**定义 6.1** 设 $(X_1,X_2,\cdots,X_n)$ 为总体 $X$ 的一个样本,$T(x_1,x_2,\cdots,x_n)$ 为不含任何未知参数的函数, 则称 $T(X_1,X_2,\cdots,X_n)$ 为一个**统计量**.

统计量是样本的函数, 但值得注意的是, 统计量中不含任何未知量. 因此, 一旦有了样本, 就可以计算出统计量. 如例 6.1 中,$\bar{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 是统计量, 但 $X-\mu$ 就不是统计量. 下面介绍一些常用的统计量.

设 $(X_1,X_2,\cdots,X_n)$ 为总体 $X$ 中抽取的一个样本. 称统计量

$$\bar{X}=\frac{1}{n}\sum_{i-1}^{n}X_i$$

为**样本均值**; 统计量

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

为**样本方差**; 需要说明的是, 这样定义的样本方差具有无偏性. 同时我们称 $S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-X)^2$ 为样本二阶中心矩. 统计量

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

为**样本标准差**; 统计量

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k, k=1,2,\cdots$$

为**样本 $k$ 阶原点矩**; 统计量

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k, k=1,2,\cdots$$

为**样本 $k$ 阶中心矩**.

**例 6.2** 设 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 为二维总体 $(X,Y)$ 的一个样本. 记

$$\begin{aligned}\bar{X}&=\frac{1}{n}\sum_{i-1}^{n}X_i, \quad S_X^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2,\\ \bar{Y}&=\frac{1}{n}\sum_{i-1}^{n}Y_i, \quad S_Y^{*2}=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2,\\ S_{XY}&=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}),\end{aligned}$$

则称 $r=S_{XY}/(S_X^*\cdot S_Y^*)$ 为样本相关系数.

前面我们讲过, 样本具有二重性. 统计量作为样本的函数, 同样具有二重性. 即在一次具体的观察中, 统计量是具体的数值; 但脱离具体的观察或试验, 统计量应看成随机变量.

统计量的分布称为**抽样分布**. 研究统计量的分布对于统计推断是十分重要的. 一般来说, 要确定一个统计量的精确分布是十分困难的, 只有在正态总体的情况下, 有比较好的结论. 这些我们将在下一节详细介绍. 对于一般情形, 我们通常利用概率论中的极限定理求出统计量的极限分布. 这在统计学中称为大样本理论.

**例 6.3** 设 $X_1,X_2,\cdots,X_n$ 为来自均值为 $\mu$ , 方差为 $\sigma^2$ 的总体的一组样本, 则当 $n$ 充分大时, 根据中心极限定理, 有

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

于是我们近似地有

$$\bar{X}\sim N(\mu,\frac{\sigma^2}{n}).$$

上例表明, 不管总体分布如何, 样本均值 $X$ 近似地服从均值为 $\mu$ , 方差为 $\frac{\sigma^2}{n}$ 的正态分布. 这在实际应用中是十分方便的.

## 6.3 正态总体

本节我们将要研究正态总体下常见统计量的精确分布. 在今后的各章中将会看到, 正态总体的研究十分重要. 这是因为在实际状况中, 正态总体是实际总体一个很好的近似. 这在概率论的极限理论的学习中我们已经体会到了. 因此, 本节的内容对于今后的学习是十分重要的.

首先我们先介绍数理统计学中占有重要地位的三大分布:$\chi^2$ 分布,$t$ 分布和 $F$ 分布.

### 6.3.1 $\chi^2$ 分布

**定义 6.2** 设 $X_1, X_2, \cdots, X_n$ 为独立同分布的随机变量, 均服从 $N(0,1)$, 则称随机变量

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2$$

为服从自由度为 $n$ 的 $\chi^2$ **分布**, 记为 $\chi_n^2 \sim \chi^2(n)$.

$\chi^2$ 分布的密度函数为

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

它的图形见图 14, 它随着 $n$ 的取值不同而不同.$\chi^2$ 分布是现代统计学的奠基人之一 Pearson 提出的, 它在统计学中有着重要的应用.

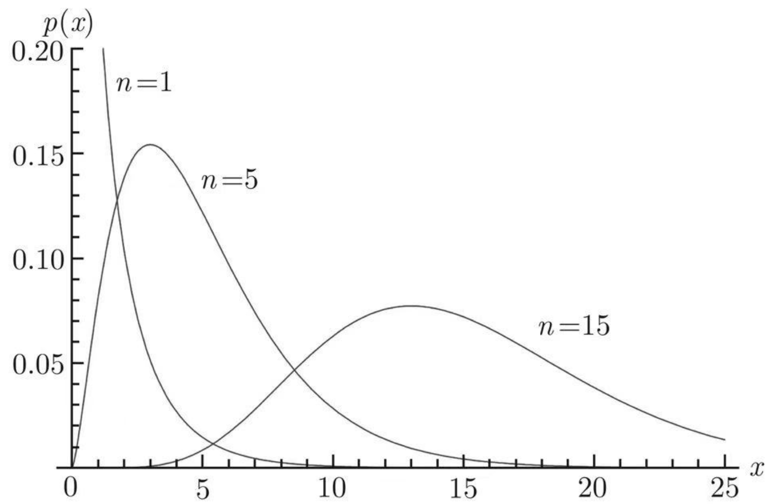


图 14

$\chi^2$ 分布具有下列性质:

**性质 1** 若 $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$, 且 $\chi_1^2$ 与 $\chi_2^2$ 相互独立, 则有 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$.

**性质 2** 若 $\chi^2 \sim \chi^2(n)$, 则 $E(X^2) = n, D(X^2) = 2n$.

性质 1 的证明由定义可以直接得到, 性质 2 的证明作为习题. 下面我们不加证明地引进一个比性质 1 更为深刻的结论.

*** 定理 6.1** 设 $X_1, X_2, \cdots, X_n$ 为独立同分布的随机变量, 均服从 $N(0,1)$. 又设

$$Q_1 + Q_2 + \cdots + Q_k = \sum_{i=1}^{n} X_i^2,$$

其中 $Q_i(i=1,2,\cdots,k)$ 是秩为 $n_i$ 的 $X_1, X_2, \cdots, X_n$ 的非负二次型, 则 $Q_i(i=1,2,\cdots,k)$ 相互独立且分别服从自由度为 $n_i$ 的 $\chi^2$ 分布的充要条件为

$$n_1 + n_2 + \cdots + n_k = n.$$

定理 6.1 称为柯赫伦 (Cochran) 分解定理, 该定理在方差分析中有重要的应用.

### 6.3.2 $t$ 分布

**定义 6.3** 设 $X \sim N(0,1), Y \sim \chi^2(n)$, 且 $X$ 与 $Y$ 相互独立, 则称随机变量

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

为服从自由度为 $n$ 的 $t$ **分布**, 记为 $T \sim t(n)$.

$t$ 分布的密度函数为

$$p(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}},$$

它的图形见图 15, 它随着 $n$ 的取值不同而不同. $t$ 分布是由英国统计学家 W.Gosset 以笔名 Student 发表的. 因此, $t$ 分布又称为学生 (Student) 分布.

如图 15 所示, $t$ 分布的密度函数是偶函数, 关于 $y$ 轴对称. 可以证明, 当 $n \to 8$ 时, $p(x)$ 收敛于标准正态分布 $\Phi(x)$. 故当 $n$ 足够大时, $t$ 分布近似 $N(0,1)$. 但 $n$ 较小时, 两者的差别较大, $t$ 分布的尾部比标准正态分布的尾部的概率大.

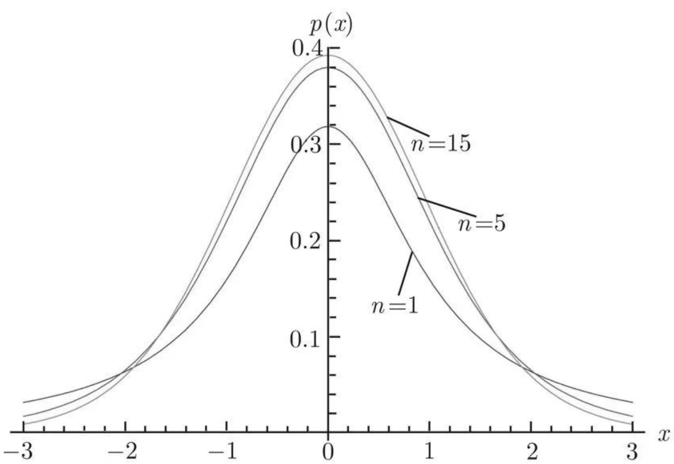


图 15

### 6.3.3 $F$ 分布

**定义 6.4** 设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, 且 $U$ 与 $V$ 相互独立, 则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

为服从自由度为 $(n_1, n_2)$ 的 $F$ **分布**, 记为 $F \sim F(n_1, n_2)$.
$F$ 分布的密度函数为

$$p(x) = \begin{cases} \frac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

它的图形见图 16.

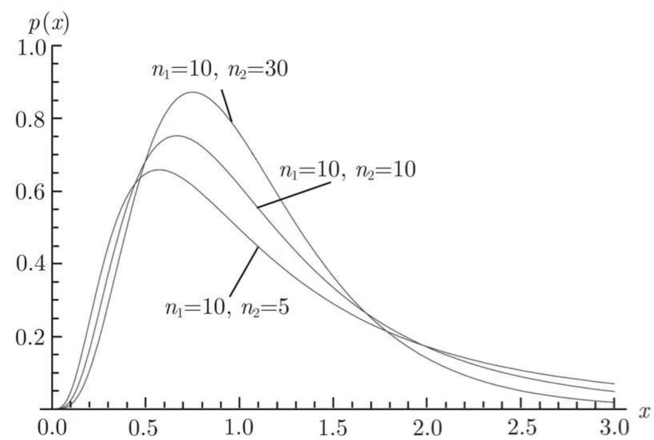


图 16

$F$ 分布是另一位现代统计学的奠基人 Fisher 提出的. 容易验证, $F$ 分布具有下列性质:
(1) 若 $F \sim F(n_1, n_2)$, 则 $\dfrac{1}{F} \sim F(n_2, n_1)$.
(2) 若 $T \sim t(n)$, 则 $T^2 \sim F(1, n)$.

### 6.3.4 上 $\alpha$ 分位点

在数理统计学中常用到上 $\alpha$ 分位点的概念.
**定义 6.5** 设 $X$ 是一个随机变量, 对于给定的数 $\alpha(0 < \alpha < 1)$, 称满足条件

$$P(X > \lambda_\alpha) = \alpha$$

的实数 $\lambda_\alpha$ 为 $X$ 的上 $\alpha$ 分位点.
上 $\alpha$ 分位点的几何意义见图 17. 其中 $p(x)$ 为 $X$ 的密度函数, 阴影部分的面积为 $\alpha$.
对于上 $\alpha$ 分位点 $\lambda_\alpha$, 当 $X \sim N(0,1)$ 时记为 $u_\alpha$; 当 $X$ 服从 $\chi^2(n), t(n), F(n_1, n_2)$ 时, 分别记为

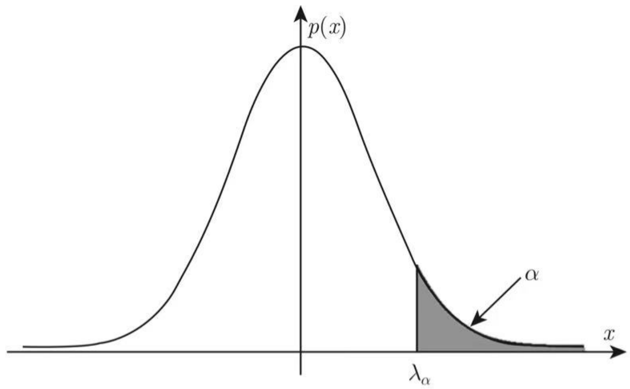


图 17

$\chi^2_\alpha(n), t_\alpha(n), F_\alpha(n_1, n_2)$. 根据这几个分布的性质, 易见

$$u_{1-\alpha} = -u_\alpha,$$
$$t_{1-\alpha}(n) = -t_\alpha(n).$$

另外, 对于 $F$ 分布的上 $\alpha$ 分位点, 有如下的性质:

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}.$$

事实上, 若 $F \sim F(n_1, n_2)$, 则

$$\begin{aligned} 1-\alpha = P[F > F_{1-\alpha}(n_1, n_2)] &= P\left[\frac{1}{F} < \frac{1}{F_\alpha(n_2, n_1)}\right] \\ &= 1 - P\left[\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\right], \end{aligned}$$

于是

$$P\left[\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right] = P\left[\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\right] = \alpha.$$

再由 $F$ 分布的性质 (1), 得到公式.

在通常的 $F$ 分布表中, 只对 $\alpha$ 比较小的值给出了上 $\alpha$ 分位点. 利用公式我们可以计算 $\alpha$ 值比较大的上 $\alpha$ 分位点.

### 6.3.5 正态总体的样本均值与样本方差的分布

对于正态总体, 样本均值与样本方差及其相关的某些统计量的分布具有非常完美的结论, 这些结论对于参数估计与假设检验具有重要的意义. 这些结果由下列定理给出.

**定理 6.2** 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的一个样本, 则

(1)$\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$;

(2)$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$;

(3)$X$ 与 $S^2$ 相互独立.

*** 证明**

(1)$X$ 是多元正态分布 $(X_1, X_2, \cdots, X_n)$ 的线性组合, 于是,$X$ 也服从正态分布. 又 $E\bar{X} = \mu, D\bar{X} = \dfrac{\sigma^2}{n}$, 因此,$X \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

再证 (2),(3). 作正交阵 $a = \begin{bmatrix} a_1 \\ \cdots \\ a_n \end{bmatrix}$, 其中 $a_1 = \left(\dfrac{1}{\sqrt{n}}, \dfrac{1}{\sqrt{n}}, \cdots, \dfrac{1}{\sqrt{n}}\right)$. 令 $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix} = A \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix} =$ $AX$. 由于 $X \sim N(\sqrt{n}\mu a_1, \sigma^2 I_n)$, 根据多元正态分布的性质,$Y \sim N(\sqrt{n}\mu A a_1, A\sigma^2 I_n A') = N(\sqrt{n}\mu A a_1, \sigma^2 I_n)$. 于是 $Y_1, Y_2, \cdots, Y_n$ 相互独立,$Y_1 = \dfrac{1}{\sqrt{n}}(X_1 + \cdots + X_n) = \sqrt{n}\bar{X}$. 且由 $A$ 的正交性, 可见 $Y_2, Y_3, \cdots, Y_n$ 的期望均为 0, 因此我们得到 $Y_i \sim N(0, \sigma^2), i = 2, 3, \cdots, n$. 同时

$$\sum_{i=1}^{n} Y_i^2 = Y'Y = (AX)'(AX) = X'A'AX = X'X = \sum_{i=1}^{n} X_i^2,$$

所以

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \\ &= \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2,\end{aligned}$$

故 $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, 且与 $X$ 相互独立. 证毕.

**推论 6.1** 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的一个样本, 则

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

**证明** 根据定理 6.2, $\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1), (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, 且相互独立. 因此,

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{(n-1)S/[(n-1)\sigma^2]}} \sim t(n-1).$$

**推论 6.2** 设 $X_1, X_2, \cdots, X_{n_1}$ 是来自正态总体 $N(\mu_1, \sigma_1^2)$ 的一个样本,$Y_1, Y_2, \cdots, Y_{n_2}$ 为来自正态总体 $N(\mu_2, \sigma_2^2)$ 的一个样本, 且两样本相互独立. 记 $\bar{X}$ 和 $\bar{Y}$ 分别为它们的样本均值, 样本方差分别为

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2,$$

$$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2.$$

则

$$F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F(n_1 - 1, n_2 - 1).$$

**证明**　根据定理 6.2 及 $F$ 分布的定义直接得到.

**推论 6.3** 设 $X_1, X_2, \cdots, X_{n_1}$ 是来自正态总体 $N(\mu_1, \sigma^2)$ 的一个样本,$Y_1, Y_2, \cdots, Y_{n2}$ 为来自正态总体 $N(\mu_2, \sigma^2)$ 的一个样本, 且两样本相互独立. 则

$$F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F(n_1 - 1, n_2 - 1).$$

**证明**　由定理 6.2, 有

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 \mu_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2\right),$$

$$[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/\sigma^2 \sim \chi^2(n_1 + n_2 - 2).$$

且两者相互独立, 再由 $t$ 分布的定义得到结论.

# 7 参数估计

参数估计与假设检验是统计推断的两个主要内容. 在很多情形下, 总体的分布类型已知, 而其中某些参数未知. 例如, 我们知道总体的分布是正态分布, 但其中的期望与方差未知. 利用样本构造合理的统计量对未知参数进行估计, 这就是参数估计的主要任务. 参数估计又可分为点估计与区间估计. 当然, 与参数估计对应的还有非参数估计, 此时, 我们连总体的分布类型也不知道, 例如, 我们只知道总体是连续性分布, 但不知道其具体类型. 如何对总体的某些特征进行估计就称为非参数估计. 本章我们只研究参数估计.

## 7.1 矩估计

设总体的分布为 $F(x;\theta)$, 其中 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)'$ 为 $k$ 维向量. 我们根据样本 $X_1,X_2,\cdots,X_n$ 构造一个统计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 作为 $\theta$ 的估计, 则称 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 $\theta$ 的**估计量**. 如果 $x_1,x_2,\cdots,x_n$ 是一组样本观察值, 代入 $\hat{\theta}$ 后得到的具体值 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 称为 $\theta$ 的**估计值**. 这样的估计称为点估计.

矩估计的思想方法是用样本矩去作为总体矩的估计. 具体地, 设参数 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)'$ 可表示为总体矩 $\mu_1,\mu_2,\cdots,\mu_k$ 的函数 $\theta_i=h_i(\mu_1,\mu_2,\cdots,\mu_k), i=1,2,\cdots,k$, 以样本矩 $A_1,A_2,\cdots,A_k$ 代替总体矩 $\mu_1,\mu_2,\cdots,\mu_k$, 所得的估计量就是**矩估计量**. 具体步骤如下:

(1) 求总体的各阶原点矩

$$\mu_i=g_i(\theta_1,\theta_2,\cdots,\theta_k), i=1,2,\cdots,k;$$

(2) 解上述方程组, 得

$$\theta_i=h_i(\mu_1,\mu_2,\cdots,\mu_k), i=1,2,\cdots,k;$$

(3) 将样本矩 $A_1,A_2,\cdots,A_k$ 代替总体矩 $\mu_1,\mu_2,\cdots,\mu_k$, 即得矩估计

$$\hat{\theta}_i=h_i(A_1,A_2,\cdots,A_k), i=1,2,\cdots,k.$$

矩估计的理论背景: 根据第五章的大数定律, 我们知道

$$A_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j\xrightarrow{P}\mu_j, j=1,2,\cdots.$$

若 $h$ 为已知的连续函数, 可以证明

$$h(A_1,A_2,\cdots,A_k)\xrightarrow{P}h(\mu_1,\mu_2,\cdots,\mu_k),$$

这称为矩估计的相合性.

**例 7.1** 设总体 $X\sim N(\mu,\sigma^2)$, 其中 $\mu,\sigma^2$ 为未知参数,$X_1,X_2,\cdots,X_n$ 是总体 $X$ 的一个样本. 求 $\mu,\sigma^2$ 的矩估计.

**解** 因为 $EX=\mu, DX=\sigma^2$, 于是

$$\mu=\mu_1;$$
$$\sigma^2=\mu_2-\mu_1^2.$$

将样本矩 $A_1, A_2$ 代入, 得到

$$\begin{aligned}\hat{\mu}=&A_1=\bar{X};\\ \hat{\sigma}^2=A_2-A_1^2=&\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.\end{aligned}$$

需要注意的是, 方差 $\sigma^2$ 的矩估计并不是样本方差 $S^2$, 而是样本二阶中心矩

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

**例 7.2** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X\sim U[\theta_1,\theta_2]$ 的一个样本, 求 $\theta_1,\theta_2$ 的矩估计.
**解** 这里有两个未知参数, 故首先求出总体的一阶, 二阶矩.

$$\begin{cases}\mu_1=EX=\dfrac{\theta_1+\theta_2}{2}\\ \mu_2=EX^2=DX+(EX)^2=\dfrac{(\theta_2-\theta_1)^2}{12}+\dfrac{(\theta_1+\theta_2)^2}{4}\end{cases}$$

解上述方程组, 得到 $\theta_1,\theta_2$ 的表达式:

$$\begin{cases}\theta_1=\mu_1-\sqrt{3(\mu_2-\mu_1^2)}\\ \theta_2=\mu_1+\sqrt{3(\mu_2-\mu_1^2)}\end{cases}$$

分别以 $A_1, A_2$ 代替 $\mu_1,\mu_2$, 即得 $\theta_1,\theta_2$ 的矩估计为

$$\begin{cases}\hat{\theta}_1=\bar{X}-\sqrt{3}S^*\\ \hat{\theta}_2=\bar{X}+\sqrt{3}S^*\end{cases}$$

**例 7.3** 设在 $n$ 次独立试验中事件 $A$ 发生 $k$ 次, 求事件 $A$ 发生的概率 $p$ 的矩估计.
**解**令 $X$ 表示在一次试验中事件 $A$ 发生与否, 即

$$X=\begin{cases}1, & 若A发生\\ 0, & 若A不发生\end{cases}$$

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X$ 的样本. 因为 $p=EX$, 故得概率 $p$ 的矩估计

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

**例 7.4** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim P(\lambda)$ 的样本, 求 $\lambda$ 的矩估计.

**解** 因为 $\lambda = EX$, 所以 $\lambda$ 的矩估计为 $\hat{\lambda} = \bar{X}$. 另外, 我们知道泊松分布的方差也是 $\lambda$, 因此也可以用样本二阶中心矩 $S^{*2} = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ 作为 $\lambda$ 的矩估计. 我们究竟采用哪一个呢? 在实际中, 我们一般选用阶数较低的样本矩, 所以, 我们选用 $X$ 作为 $\lambda$ 的矩估计.

**例 7.5** 设 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ 为二维总体 $(X, Y)$ 的一个样本, 求相关系数 $\rho$ 的矩估计.

**解** 因为 $\rho = \dfrac{\operatorname{cov}(X,Y)}{\sqrt{DX \cdot DY}} = EXY - EX \cdot EY[EX^2 - (EX)^2] \cdot [EY^2 - (EY)^2]$, 故相关系数 $\rho$ 是矩 $(EX, EX^2, EY, EY^2, EXY)$ 的函数. 将样本矩代替原点矩, 得到 $\rho$ 的矩估计为

$$\hat{\rho} = \frac{\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i Y_i - \bar{X} \cdot \bar{Y}}{\sqrt{\left(\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - \bar{X}^2\right)\left(\dfrac{1}{n}\sum\limits_{i=1}^{n} Y_i^2 - \bar{Y}^2\right)}} = \frac{S_{XY}}{S_X^* \cdot S_Y^*}.$$

其中

$$\begin{aligned}
\bar{X} &= \frac{1}{n}\sum_{i-1}^{n} X_i, \quad S_X^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \\
\bar{Y} &= \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad S_Y^{*2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \\
S_{XY} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).
\end{aligned}$$

事实上, 由例 6.2 可知,$\rho$ 的矩估计就是样本相关系数 $r$.

矩估计是一种直观, 简便的估计方法, 它的适用范围很广, 并不需要知道总体分布的具体类型, 只需计算总体矩的表达式即可. 但由于矩估计方法没有充分利用总体分布 $F(x;\theta)$ 中的信息, 因此, 矩估计往往精度不高.

## 7.2 极大似然估计

极大似然估计是英国著名统计学家 Fisher 于 1922 年提出的. 迄今为止, 该方法仍然是统计学中最重要的方法之一.

极大似然估计的思想是非常直观的. 例如, 靶场上打了一枪, 发现这一枪打中了 10 环. 现可能有两个射手打了这一枪, 其中一人是神射手, 另一人是普通射手. 问这一枪可能是谁打的? 这就相当于我们有了一个样本 (10 环), 两个射手相当于两个参数, 问题是选取哪一个参数作为估计? 从直观上, 我们肯定会选取神射手 (因为神射手打中 10 环的概率大). 这就是极大似然估计的思想. 下面我们再通过一个例子说明极大似然估计方法的思想.

在已有样本的基础上, 要选择参数的一个合理的估计值, 就是要使得参数在取该估计值时样本发生的可能性达到最大. 这就是极大似然估计的思想方法.

一般地, 设总体 $X$ 为连续型随机变量, 其密度函数为 $p(x;\theta), \theta \in \Theta$,$\theta$ 为待估参数,$\Theta$ 为参数 $\theta$ 的取

值范围.$(X_1,X_2,\cdots,X_n)$ 为来自总体 $X$ 的样本, 则 $(X_1,X_2,\cdots,X_n)$ 的概率密度函数为 $\prod\limits_{i=1}^{n}p(x_i;\theta)$. 又设 $(x_1,x_2,\cdots,x_n)$ 是样本的一组观察值, 那么样本 $(X_1,X_2,\cdots,X_n)$ 落在点 $(x_1,x_2,\cdots,x_n)$ 的邻域内的概率近似为 $\prod\limits_{i=1}^{n}p(x_i;\theta)\mathrm{d}x_i$.

若总体 $X$ 为离散型随机变量, 其概率函数为 $p(x;\theta),\theta\in\Theta,\theta$ 为待估参数,$\Theta$ 为参数 $\theta$ 的取值范围. 同样设 $(X_1,X_2,\cdots,X_n)$ 为来自总体 $X$ 的样本, 则 $(X_1,X_2,\cdots,X_n)$ 的联合概率仍为 $\prod\limits_{i=1}^{n}p(x_i;\theta)$.

极大似然估计法就是选取使得上面的概率达到最大的参数值 $\hat{\theta}$ 作为 $\theta$ 的估计. 为方便起见, 不论总体 $X$ 为离散型随机变量或为连续型随机变量, 均记

$$L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta).$$

当样本值 $(x_1,x_2,\cdots,x_n)$ 取定时,$L(x_1,x_2,\cdots,x_n;\theta)$ 是关于 $\theta$ 的函数, 称之为**似然函数**. 满足下式

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n;\theta).$$

的最大值点 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 称为 $\theta$ 的**极大似然估计值**,$\hat{\theta}(X_1,X_2,\cdots,X_n)$ 称为**极大似然估计量**.

由于 $L(\theta)$ 是连乘的形式, 求最大值并不方便. 注意到 $\ln L(\theta)$ 是 $L(\theta)$ 的单调函数, 因而有相同的最大值点. 在很多情况下, 求 $\ln L(\theta)$ 的最大值点比较简单. 我们称 $\ln L(\theta)$ 为**对数似然函数**, 称

$$\frac{\partial\ln L(\theta)}{\partial\theta}=\sum_{i=1}^{n}\frac{\partial\ln p(x_i;\theta)}{\partial\theta}=0.$$

为**似然方程 (组)**. 解似然方程 (组), 得到的解往往就是极大似然估计. 下面举例说明求极大似然估计的方法.

**例 7.7** 设总体 $X\sim B(1,p)$, $X_1,X_2,\cdots,X_n$ 是来自总体 $X$ 的样本, 试求参数 $p$ 的极大似然估计.

**解** 设 $(x_1,x_2,\cdots,x_n)$ 是样本 $(X_1,X_2,\cdots,X_n)$ 的一组观察值.$X$ 的分布律为

$$P(X=x)=p^x(1-p)^{1-x},x=0,1.$$

所以似然函数为

$$L(p)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}.$$

故对数似然函数为

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

似然方程如下:

$$\frac{\mathrm{d}\ln L(p)}{\mathrm{d}p}=\frac{\sum\limits_{i=1}^{n}x_i}{p}-\frac{n-\sum\limits_{i=1}^{n}x_i}{1-p}=0.$$

解得 $p$ 的极大似然估计值为

$$\hat{p}=\frac{\sum\limits_{i=1}^{n}x_i}{n}=\bar{x}.$$

$p$ 的极大似然估计量为

$$\hat{p}=\frac{\sum\limits_{i=1}^{n}X_i}{n}=\bar{X}.$$

**例 7.8** 设总体 $X\sim P(\lambda)$,$X_1,X_2,\cdots,X_n$ 是来自总体 $X$ 的样本, 试求参数 $\lambda$ 的极大似然估计.

**解** 设 $(x_1,x_2,\cdots,x_n)$ 是样本 $(X_1,X_2,\cdots,X_n)$ 的一组观察值.$X$ 的分布律为

$$P(X=x)=\frac{\lambda^x}{x!}e^{-\lambda},x=0,1,2,\cdots.$$

于是参数 $\lambda$ 的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda\sum\limits_{i=1}^{n}x_i}{\prod\limits_{i=1}^{n}x_i!}e^{-n\lambda}.$$

似然方程为

$$\frac{\mathrm{d}\ln L(\lambda)}{\mathrm{d}\lambda}=\frac{\sum\limits_{i=1}^{n}x_i}{\lambda}-n=0.$$

解得 $\lambda$ 的极大似然估计值为

$$\hat{\lambda}=\frac{\sum\limits_{i=1}^{n}x_i}{n}.$$

$\lambda$ 的极大似然估计量为

$$\hat{\lambda}=\frac{\sum\limits_{i=1}^{n}X_i}{n}.$$

**例 7.9** 设总体 $X\sim N(\mu,\sigma^2)$, 其中 $\mu,\sigma^2$ 为未知参数,$X_1,X_2,\cdots,X_n$ 是 $X$ 的一个样本. 求 $\mu,\sigma^2$ 的极大似然估计.

**解**　似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$
$$=(2\pi)^{-\frac{n}{2}}(\sigma^2)-\frac{n}{2}\exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right]$$

它的对数为

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

似然方程组为

$$\begin{cases}\dfrac{\partial\ln\ L}{\partial\mu}=\dfrac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}=0\\ \dfrac{\partial\ln\ L}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0\end{cases}$$

由前一式解得 $\hat{\mu}=\bar{x}$, 代入后一式得 $\hat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$. 因此,$\mu,\sigma^2$ 的极大似然估计量分别为 $\bar{X},S^{*2}=\frac{n-1}{n}S^2$. 它们与矩估计相同.

**例 7.10** 设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim U[\theta_1,\theta_2]$ 的样本, 求 $\theta_1,\theta_2$ 的极大似然估计.

**解**　$X$ 的密度函数为

$$p(x;\theta_1,\theta_2)=\begin{cases}\dfrac{1}{\theta_2-\theta_1}, & \theta_1\le x\le\theta_2\\ 0, & \text{其他}\end{cases}$$

似然函数为

$$L(\theta_1,\theta_2)=\begin{cases}\dfrac{1}{(\theta_2-\theta_1)^n}, & \theta_1\le x_i\le\theta_2, i=1,2,\cdots,n\\ 0, & \text{其他}\end{cases}$$

显然,$L(\theta_1,\theta_2)$ 作为 $\theta_1,\theta_2$ 的函数是不连续的, 此时我们不能用似然方程组来解极大似然估计, 而必须从极大似然估计的定义出发, 求 $L(\theta_1,\theta_2)$ 的最大值. 为使 $L(\theta_1,\theta_2)$ 达到最大, 须使 $\theta_2-\theta_1$ 尽可能小. 但同时注意到 $\theta_2$ 不能小于 $\max(x_1,x_2,\cdots,x_n)$,$\theta_1$ 不能大于 $\min(x_1,x_2,\cdots,x_n)$, 否则 $L(\theta_1,\theta_2)=0$. 故得 $\theta_1,\theta_2$ 的极大似然估计为

$$\hat{\theta_1}=\min(x_1,x_2,\cdots,x_n),$$
$$\hat{\theta_2}=\max(x_1,x_2,\cdots,x_n).$$

由于考虑到总体的分布, 极大似然估计通常比矩估计优良. 但极大似然估计的计算较复杂, 往往需要计算机才能得到近似解. 另外, 极大似然估计还具有一个优良的性质: **不变性原则**. 设 $\hat{\theta}$ 是参数 $\theta$ 的极

大似然估计,$\varphi(\theta)$ 有单值反函数, 则 $\varphi(\hat{\theta})$ 是 $\varphi(\theta)$ 的极大似然估计, 即

$$\widehat{\varphi(\theta)} = \varphi(\hat{\theta}).$$

例如, 在例 7.9 中, 我们求得 $\sigma^2$ 的极大似然估计为 $S^{*2}$, 因 $\varphi(\sigma) = \sigma^2(\sigma \geq 0)$ 有单值反函数, 由此性质可得 $\sigma$ 的极大似然估计为 $S^*$.

## 7.3 估计量的评价标准

通过上面两节的内容可以看出, 由不同的方法可以得到不同的参数估计. 那么, 究竟采用哪一个估计量呢? 这就涉及估计量的评价标准. 这里只介绍常用的三种标准, 无偏性, 均方误差准则和一致性.

### 7.3.1 无偏性

**定义 7.1** 设 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 为参数 $\theta$ 的一个估计量,$\Theta$ 为参数 $\theta$ 的取值范围, 若对任意 $\theta \in \Theta$, 有

$$E[\hat{\theta}(X_1, X_2, \cdots, X_n)] = \theta,$$

则称 $\hat{\theta}$ 是 $\theta$ 的**无偏估计量**.

无偏性的实际意义是, 用估计量 $\hat{\theta}$ 对未知参数 $\theta$ 进行估计, 有时会高于 $\theta$, 有时会低于 $\theta$, 但平均来说它等于未知参数 $\theta$. 也就是说, 没有系统误差.

**例 7.11** 设总体 $X$ 的 $k$ 阶原点矩 $\mu_k$ 存在,$X_1, X_2, \cdots, X_n$ 是来自总体的一个样本. 则 $k$ 阶样本原点矩 $A_k$ 是 $\mu_k$ 的无偏估计.

**证明** $E(A_k) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n}\sum_{i=1}^{n} EX_i^k = \mu_k.$

特别地,$\bar{X}$ 是总体的数学期望 $EX$ 的无偏估计.

**例 7.12** 设总体 $X$ 的方差 $\sigma^2$ 存在且有限, $X_1, X_2, \cdots, X_n$ 是来自总体的一个样本. 则样本方差 $S^2$ 是方差 $\sigma^2$ 的无偏估计.

**证明** 因为

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

故得

$$\begin{aligned} E(S^2) &= \frac{1}{n-1}\sum_{i=1}^{n} E(X_i^2) - \frac{n}{n-1}E(\bar{X})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(\sigma^2 + \mu^2) - \frac{n}{n-1}(\frac{\sigma^2}{n} + \mu^2) = \sigma^2. \end{aligned}$$

同时可见, 样本二阶中心矩 $S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 不是 $\sigma^2$ 的无偏估计. 但有 $E(S^{*2}) = \frac{n-1}{n}\sigma^2 \to \sigma^2$,, 此时我们称 $S^{*2}$ 为 $\sigma^2$ 的**渐进无偏估计**.

**例 7.13** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim U[0, \theta]$ 的样本, 令 $X_{(n)} = max(X_1, X_2, \cdots, X_n)$. 试证:$\hat{\theta} = \frac{n+1}{n}X_{(n)}$ 是 $\theta$ 的无偏估计.

**证明** 总体 $X$ 的分布函数为

$$F(x)=\begin{cases}0, & x\le 0\\ \dfrac{x}{\theta}, & 0<x\le\theta\\ 1, & x>\theta\end{cases}$$

于是 $X_{(n)}$ 的分布函数为 $F^n(x)$, 所以 $X_{(n)}$ 的密度函数为

$$p_n(x;\theta)=\begin{cases}\dfrac{nx^{n-1}}{\theta^n}, & 0\le x\le\theta\\ 0, & \text{其他}\end{cases}$$

故得

$$\begin{aligned}E(\hat{\theta})&=\frac{n+1}{n}E(X_{(n)})\\&=\frac{n+1}{n}\int_0^\theta\frac{nx^n}{\theta^n}\mathrm{d}x\\&=\frac{n+1}{\theta^n}\frac{x^{n+1}}{n+1}\bigg|_0^\theta=\theta.\end{aligned}$$

所以 $\hat{\theta}=\dfrac{n+1}{n}X_{(n)}$ 是 $\theta$ 的无偏估计.

注意, 即使 $E(\hat{\theta})=\theta$, 不一定有 $E[g(\hat{\theta})]=g(\theta)$, 其中 $g(\theta)$ 是 $\theta$ 的函数. 这说明, 当 $\theta$ 是 $\theta$ 的无偏估计, 未必有 $g(\hat{\theta})$ 是 $g(\theta)$ 的无偏估计.

### 7.3.2 均方误差准则

作为参数 $\theta$ 的估计 $\hat{\theta}(X_1,X_2,\cdots,X_n)$, 我们当然希望它们的差距 $\hat{\theta}-\theta$ 越小越好, 由于 $\hat{\theta}$ 是随机变量, 我们用 $E(\hat{\theta}-\theta)^2$ 来衡量估计量 $\hat{\theta}$ 的好坏, 并将其称为均方误差, 记为 $M(\hat{\theta},\theta)$, 即

$$M(\hat{\theta},\theta)=E(\hat{\theta}-\theta)^2.$$

显然, 均方误差越小越好, 这一准则称为**均方误差准则**. 又

$$\begin{aligned}M(\hat{\theta},\theta)&=E(\hat{\theta}-E\hat{\theta}+E\hat{\theta}-\theta)^2\\&=E(\hat{\theta}-E\hat{\theta})^2+2E(\hat{\theta}-E\hat{\theta})(E\hat{\theta}-\theta)+(E\hat{\theta}-\theta)^2\\&=D(\hat{\theta})+(E\hat{\theta}-\theta)^2,\end{aligned}$$

因此, 均方误差可以分解为两部分

$$M(\hat{\theta},\theta)=D(\hat{\theta})+(E\hat{\theta}-\theta)^2,$$

第一部分是 $\hat{\theta}$ 的方差, 第二部分是估计量偏差 $E\hat{\theta}-\theta$ 的平方. 如果估计量是无偏估计, 那么第二部分为零, 均方误差就变为方差.

**例 7.14** 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim U[0,\theta]$ 的样本, 试比较 $\hat{\theta}_1 = 2\bar{X}$ 与 $\hat{\theta}_2 = X_{(n)}$ 的均方误差.

**解**因为 $\hat{\theta}_1$ 是无偏估计, 所以

$$M(\hat{\theta}_1, \theta) = D(\hat{\theta}_1) = 4D(\bar{X}) = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}.$$

又由例 7.13 知 $X_{(n)}$ 的密度函数如下:

$$p_n(x;\theta) = \begin{cases} \dfrac{nx^{n-1}}{\theta^n}, & 0 \le x \le \theta \\ 0, & \text{其他} \end{cases}$$

于是 $E(X_{(n)}) = \dfrac{n}{n+1}\theta$, 且

$$\begin{aligned} E(X_{(n)}^2) &= \int_0^\theta \frac{nx^{n+1}}{\theta^n} \mathrm{d}x = \frac{n}{n+2} \frac{x^{n+2}}{\theta^n}\bigg|_0^\theta \\ &= \frac{n}{n+2}\theta^2. \end{aligned}$$

所以

$$\begin{aligned} M(\hat{\theta}_2, \theta) &= D(X_{(n)}) + (EX_{(n)} - \theta)^2 \\ &= \frac{n\theta^2}{(n+2)(n+1)^2} + \frac{\theta^2}{(n+1)^2} \\ &= \frac{2\theta^2}{(n+2)(n+1)}. \end{aligned}$$

当 $n > 2$ 时, 有 $M(\hat{\theta}_1, \theta) > M(\hat{\theta}_2, \theta)$. 事实上, 容易看出,$\hat{\theta}_1$ 是矩估计, 而 $\hat{\theta}_2$ 则是极大似然估计. 这说明极大似然估计优于矩估计.

### 7.3.3 * 一致性

无偏性与均方误差都是在样本量 $n$ 固定的前提下对估计量进行研究, 有时我们也可以考虑当样本量 $n \to \infty$ 时估计量的性质. 从直观上, 当样本量越来越多时, 样本中含有的关于未知参数的信息也越来越多, 因此估计也应该越准确. 也就是说, 估计量应越来越接近于真实参数. 这就是一致性的概念.

**定义 7.2** 设 $\hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 为参数 $\theta$ 的一个估计量,$\Theta$ 为参数 $\theta$ 的取值范围, 若对任意 $\theta \in \Theta$, 有

$$\hat{\theta_n} \xrightarrow{P} \theta.$$

则称 $\hat{\theta}_n$ 是 $\theta$ 的**一致估计量**. 一致性是在样本量趋于无穷大时对估计量的要求, 因此, 它是估计量的大样本性质, 也称为**相合性**. 由本章第一节知, 矩估计往往具有相合性.

## 7.4 区间估计

### 7.4.1 基本概念与枢轴变量法

前面我们讨论了参数估计中的点估计问题. 除了用一个点去估计未知参数外, 我们还可以构造一个区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 来估计参数 $\theta$ 的范围. 由于 $\hat{\theta}_1, \hat{\theta}_2$ 均为随机变量, 因此, 我们还必须指出该区间以多大概率包含未知参数 $\theta$, 这称为区间的置信度.

**定义 7.3** 设 $\theta$ 是总体 $X$ 的未知参数,$X_1, X_2, \cdots, X_n$ 是来自总体 $X$ 的样本. 若对事先给定的常数 $\alpha(0 < \alpha < 1)$, 存在两个统计量 $\hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 与 $\hat{\theta}_2(X_1, X_2, \cdots, X_n)$, 使得

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha,$$

则称区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 是 $\theta$ 的置信度为 $1 - \alpha$ 的**置信区间**,$\hat{\theta}_1, \hat{\theta}_2$ 分别称为**置信下限**和**置信上限**,$1 - \alpha$ 称为**置信度**或**置信系数**.

应该特别指出的是, 上式的含义是在重复取得多组样本的情况下, 将得到很多不同的区间 $(\hat{\theta}_1, \hat{\theta}_2)$, 这些区间中大约有 $100(1-\alpha)\%$ 的区间包含未知参数 $\theta$. 但对于一组样本来说, 绝不能认为“$\theta \in (\hat{\theta}_1(x_1, x_2, \cdots, x_n), \hat{\theta}_2(x_1, x_2, \cdots, x_n))$ 的概率为 $1-\alpha$”. 因为对于一组样本观察值来说,$\hat{\theta}_1(x_1, x_2, \cdots, x_n)$ ,$\hat{\theta}_2(x_1, x_2, \cdots, x_n)$ 均为具体的数值, 此时要么区间 $(\hat{\theta}_1(x_1, x_2, \cdots, x_n), \hat{\theta}_2(x_1, x_2, \cdots, x_n))$ 包含未知参数 $\theta$, 要么该区间不包含未知参数 $\theta$.

另外, 置信区间的长度可以看成区间估计的精度. 当然我们希望区间估计既有高的置信度, 又有好的精度. 但一般来说, 精度与置信度二者之间是矛盾的. 随机区间的长度越长, 置信度就越高, 但精度下降; 反之, 随机区间的长度越短, 精度提高, 但置信度下降. 在实际问题中, 我们总是在保证置信度的条件下, 尽可能地提高精度.

在实际问题中, 有时我们只关心未知参数的上限或下限. 例如, 工厂生产一批电子产品, 需要估计产品的平均寿命. 产品寿命越长越好, 因此我们只关心产品寿命的下限. 这样就需要引进单侧置信区间的概念.

**定义 7.4** 设 $\theta$ 是总体 $X$ 的未知参数,$X_1, X_2, \cdots, X_n$ 是来自总体 $X$ 的样本. 若对事先给定的常数 $\alpha(0 < \alpha < 1)$, 存在一个统计量 $\hat{\theta}_1$, 使得

$$P(\hat{\theta}_1 < \theta) = 1 - \alpha,$$

则称区间 $(\hat{\theta}_1, +\infty)$ 是 $\theta$ 的置信度为 $1 - \alpha$ 的单侧置信区间,$\hat{\theta}_1$ 称为置信度为 $1 - \alpha$ 的单侧置信下限. 若存在统计量 $\hat{\theta}_2$, 使得

$$P(\theta < \hat{\theta}_2) = 1 - \alpha,$$

则称区间 $(-\infty, \hat{\theta}_2)$ 是 $\theta$ 的置信度为 $1 - \alpha$ 的**单侧置信区间**,$\hat{\theta}_2$ 称为置信度为 $1 - \alpha$ 的**单侧置信上限**.

下面通过一个例子说明如何求区间估计.

**例 7.15** 设一批螺栓的直径服从正态分布 $N(\mu, 0.1^2)$, 现从中抽取 25 枚, 测得其平均直径为 2.024 厘米. 求螺栓的直径 $\mu$ 的置信度为 95% 的置信区间.

**解** 首先样本均值 $X$ 是总体均值 $\mu$ 的一个合理估计 (即是矩估计, 又是极大似然估计). 由此构造一个样本函数

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

该函数含有未知参数 $\mu$, 但它的分布不含任何未知参数. 对于给定的置信度为 $1-\alpha$, 有

$$P(|U|<u_{\alpha/2})=1-\alpha,$$

其中 $u_{\alpha/2}$ 为标准正态分布的上 $\dfrac{\alpha}{2}$ 分位点. 利用不等式变形可得

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

即得

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

于是我们求得 $\mu$ 的置信度为 $1-\alpha$ 的置信区间为 $\left(\bar{X}-u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}},\bar{X}+u_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$.
将 $\sigma=0.1,n=25,\alpha=0.05,u_{\alpha/2}=u_{0.025}=1.96,\bar{X}=2.024$ 代入, 得

$$\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}}=2.024-1.96\times\frac{0.1}{\sqrt{25}}=1.9848,\bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}}=2.024+1.96\times\frac{0.1}{\sqrt{25}}=2.0632.$$

所以 $\mu$ 的置信度为 95% 的置信区间为 $(1.9848,2.0632)$.

总结例 7.15, 可以得到求区间估计的一般方法, 具体步骤如下:

(1) 先找一个样本函数 $U(X_1,X_2,\cdots,X_n;\theta)$. 它包含待估参数 $\theta$, 而不包含其他未知参数, 且 $U$ 的分布已知, 不依赖于任何未知参数. 这样的函数称为枢轴变量 (注意枢轴变量不是统计量, 因为它含有未知参数).

(2) 对事先给定的置信度为 $1-\alpha$, 根据 $U$ 的分布找到两个常数 $a,b$, 使得

$$P(a<U<b)=1-\alpha.$$

(3) 利用不等式变形, 由 $a<U<b$ 解出 $\hat{\theta}_1<\theta<\hat{\theta}_2$, 则 $(\hat{\theta}_1,\hat{\theta}_2)$ 就是 $\theta$ 的置信度为 $1-\alpha$ 的置信区间.

这种方法称为**枢轴变量法**. 其中构造枢轴变量 $U(X_1,X_2,\cdots,X_n;\theta)$ 是其关键, 它的构造往往是通过点估计而来的. 单侧置信区间的求法是类似的, 在此我们不赘述了.

### 7.4.2 正态总体 $N(\mu,\sigma^2)$ 中均值 $\mu$ 的置信区间

设给定置信度为 $1-\alpha,X_1,X_2,\cdots,X_n$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一组样本.

(1) $\sigma^2$ 已知, 求 $\mu$ 的置信区间.

由例 7.15, 取枢轴变量

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

再由 $P(|U|<u_{\alpha/2})=1-\alpha$, 利用不等式变形解得

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

于是,$\mu$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\bar{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}},\bar{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

这里顺便讲一下单侧置信区间的求法. 以置信上限为例, 枢轴变量仍为

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

再由 $P(U>-u_\alpha)=1-\alpha$, 利用不等式变形解得

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

于是,$\mu$ 的置信度为 $1-\alpha$ 的单侧置信区间为 $\left(-\infty,\bar{X}+u_\alpha\frac{\sigma}{\sqrt{n}}\right)$. 类似地,$\mu$ 的置信度为 $1-\alpha$ 的另外一个单侧置信区间为 $\left(\bar{X}-u_\alpha\frac{\sigma}{\sqrt{n}},+\infty\right)$.

(2)$\sigma^2$ 未知, 求 $\mu$ 的置信区间.

由推论 6.1, 取

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

它包含待估参数 $\mu$, 且分布完全确定, 不依赖任何未知参数. 在给定的置信度为 $1-\alpha$ 下,$P[|T|<t_{\alpha/2}(n-1)]=1-\alpha$. 利用不等式变形解得

$$P\left[\bar{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}<\mu<\bar{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right]=1-\alpha.$$

于是,$\mu$ 的置信度为 $1-\alpha$ 的置信区间为

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

比较两个公式可以发现, 当 $\sigma^2$ 未知时, 我们只需将 $\sigma$ 换成样本标准差 $S$(它是 $\sigma$ 的合理估计), 当然区间估计中的上 $\alpha$ 分位点 $u_{\alpha/2}$ 也要换成 $t_{\alpha/2}(n-1)$.

**例 7.16** 设有一批糖果, 现随机抽取 12 包, 称得质量 (单位: 千克) 为

| | | | | | |
|---|---|---|---|---|---|
| 9.9 | 10.1 | 10.3 | 10.4 | 10.5 | 10.2 |
| 9.7 | 9.8 | 10.1 | 10.0 | 9.8 | 10.3 |

假定质量服从正态分布, 试求总体均值的置信度为 95% 的置信区间.

**解** 由样本观察值算得 $\bar{X}=10.092,S=0.2575,t_{0.025}(11)=2.201$, 代入上式

$$\bar{X}-t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}=10.092-2.201\times\frac{0.2575}{\sqrt{12}}=9.9284,$$
$$\bar{X}+t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}=10.092+2.201\times\frac{0.2575}{\sqrt{12}}=10.2556,$$

所以均值的置信度为 95% 的置信区间为 (9.9284,10.2556).

### 7.4.3 正态总体 $N(\mu,\sigma^2)$ 中方差 $\sigma^2$ 的置信区间

这里只介绍 $\mu$ 未知的情形.

由定理 6.2 知

$$(n-1)S^2/\sigma^2 \sim \chi^2(n-1),$$

上式分布已知, 不依赖于任何未知参数. 取作枢轴变量. 故得

$$P\left[\chi^2_{1-\alpha/2}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}(n-1)\right] = 1-\alpha.$$

解得

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}(n-1)\right] = 1-\alpha.$$

于是,$\sigma^2$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

注意, 在密度函数不对称时, 如 $\chi^2$ 分布和 $F$ 分布, 我们仍然用对称的分位点来确定置信区间. 这样得到的置信区间的长度未必是最短的, 但计算方便. 如要求最短的置信区间, 则计算过于复杂, 实际中一般不会去求.

**例 7.17** 设将一个物体称了 10 次, 得到的质量 (单位: 千克) 为

| | | | | |
|---|---|---|---|---|
| 10.1 | 10 | 9.8 | 10.5 | 9.7 |
| 10.1 | 9.9 | 10.2 | 10.3 | 9.9 |

表 8

假定质量服从正态分布, 试求总体方差 $\sigma^2$ 的置信度为 95% 的置信区间.

**解** 由样本观察值算得 $S^2 = 0.0583$, 又 $n = 10, x^2_{0.025}(9) = 19.023, x^2_{0.975}(9) = 2.70$. 代入上式得 2 的置信度为 95% 的置信区间为 (0.028,0.194).

### 7.4.4 两个正态总体 $N(\mu_1,\sigma_1^2), N(\mu_2,\sigma_2^2)$ 的均值差 $\mu_1-\mu_2$ 的置信区间

两个正态总体的区间估计问题, 在实际应用中经常遇到. 例如, 比较甲, 乙两个工厂生产的药品的治疗效果, 考察一项新技术对产品的质量指标的作用等. 这些问题通常可以归结为研究两个正态总体的均值差的区间估计.

设 $X_1, X_2, \cdots, X_{n_1}$ 是来自正态总体 $N(\mu_1,\sigma_1^2)$ 的一个样本,$Y_1, Y_2, \cdots, Y_{n_2}$ 为来自正态总体 $N(\mu_2,\sigma_2^2)$ 的一个样本, 且两样本相互独立. 记 $X$ 和 $Y$ 分别为它们的样本均值,$S_1^2$ 和 $S_2^2$ 分别为它们的样本方差.

(1)$\sigma_1^2, \sigma_2^2$ 均已知, 求 $\mu_1-\mu_2$ 的置信区间.

取

$$U=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1),$$

再由 $P(|U|<u_{\alpha/2})=1-\alpha$ 利用不等式变形解得

$$P\left(\bar{X}-\bar{Y}-u_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}<\mu_1-\mu_2<\bar{X}-\bar{Y}+u_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)=1-\alpha.$$

即得 $\mu_1-\mu_2$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\bar{X}-\bar{Y}-u_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\bar{X}-\bar{Y}+u_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right).$$

(2)$\sigma_1^2=\sigma_2^2=\sigma^2$, 但 $\sigma^2$ 未知, 求 $\mu_1-\mu_2$ 的置信区间.
由推论 6.3, 取

$$T=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}\sim t(n_1+n_2-2),$$

在给定的置信度为 $1-\alpha$ 下,$P[|T|<t_{\alpha/2}(n_1+n_2-2)]=1-\alpha$. 利用不等式变形解得

$$\begin{aligned}P\Bigg[\bar{X}-\bar{Y}-t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2\\+t_\alpha/2(n_1+n_2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\Bigg]=1-\alpha,\end{aligned}$$

其中 $S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$. 从而 $\mu_1-\mu_2$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\bar{X}-\bar{Y}-t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\bar{X}-\bar{Y}+t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right).$$

**例 7.18** 设某工厂利用两条自动化流水线灌装矿泉水. 随机地抽取样本 $X_1,X_2,\cdots,X_{12}$ 和 $Y_1,Y_2,\cdots,Y_{17}$, 它们是每瓶矿泉水的体积 (单位: 毫升). 计算得样本均值 $\bar{X}=501.1$ 和 $\bar{Y}=499.7$, 样本方差 $S_1^2=2.4,S_2^2=4.7$. 假定两条流水线生产的矿泉水体积均服从正态分布, 且它们的方差相等. 求 $\mu_1-\mu_2$ 的置信度为 95% 的置信区间.

**解** 首先计算 $S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}=\sqrt{\frac{11\times2.4+16\times4.7}{12+17-2}}=1.94$. 又 $t_{0.025}(27)=2.05$, 代入公式,$\mu_1-\mu_2$ 的置信度为 95% 的置信区间为 (-0.101,2.901).

在例 7.18 中,$\mu_1-\mu_2$ 的区间估计包含零, 这时我们可以认为 $\mu_1$ 与 $\mu_2$ 没有显著差异. 详细讨论见下章内容-假设检验.

### 7.4.5 两个正态总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ 的方差比 $\sigma_1^2/\sigma_2^2$ 的置信区间

这里我们只介绍 $\mu_1, \mu_2$ 未知的情形. 记号同上一小节. 由推论 6.2, 取

$$F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F(n_1 - 1, n_2 - 1).$$

$F$ 的分布已知, 不依赖于任何未知参数. 将 $F$ 取作枢轴变量. 在给定的置信度为 $1-\alpha$ 下得

$$P[F_{1-\alpha/2}(n_1-1, n_2-1) < \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} < F_{\alpha/2}(n_1-1, n_2-1)] = 1-\alpha.$$

利用不等式变形, 得到 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left( \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\alpha/2}(n_1-1, n_2-1)} \right).$$

**例 7.19** 设两个化验员独立地对某种聚合物的含氯量分别做了 10 次和 11 次测定, 算得样本方差为 $S_1^2 = 0.5419, S_2^2 = 0.6065$. 假定测量值均服从正态分布. 求方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信度为 90% 的置信区间.

**解** 按题意 $n_1 = 10, n_2 = 11, 1-\alpha = 90\%$. 于是得 $F_{0.05}(9, 10) = 3.02$ 及 $F_{0.95}(9, 10) = \dfrac{1}{F_{0.05(10,9)}} = \dfrac{1}{3.14}$, 从而由公式得 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信度为 90% 的置信区间为

$$\left( \frac{0.5419}{0.6065} \times \frac{1}{3.02}, \frac{0.5419}{0.6065} \times 3.14 \right) = (0.296, 2.806).$$

### 7.4.6 * 非正态总体均值的区间估计 (大样本法)

当总体分布非正态时, 一般很难求出统计量的具体分布. 此时我们采取大样本方法. 在样本量较大的情况下, 利用极限定理求出枢轴变量的近似分布, 然后求得参数的区间估计.

设 $X_1, X_2, \cdots, X_n$ 为来自均值为 $\mu$, 方差为 $\sigma^2$ 的总体的一组样本, 给定置信度为 $1-\alpha$, 求均值 $\mu$ 的区间估计.

当 $n$ 充分大时, 根据中心极限定理, 有

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

于是

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

若 $\sigma$ 未知, 可以用样本标准差 $S$ 代替, 得

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

将其取作枢轴变量, 再由 $P(|U| < u_{\alpha/2}) = 1-\alpha$, 利用不等式变形解得均值 $\mu$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\bar{X} - u_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + u_{\alpha/2}\frac{S}{\sqrt{n}}\right).$$

**例 7.20** 设总体服从 0-1 分布,$X_1, X_2, \cdots, X_n$ 为来自总体的一组样本. 求 $p$ 的置信度为 $1-\alpha$ 的置信区间.

**解** 由于 $p = EX$, 注意到

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

可以用公式直接计算得到 $p$ 的置信度为 $1-\alpha$ 的置信区间如下:

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

我们这里再介绍另外一种求 $p$ 的置信度为 $1-\alpha$ 的置信区间的方法. 因为总体 $X \sim B(1,p)$, 所以 $\sum\limits_{i=1}^{n} X_i \sim B(n,p)$. 依中心极限定理, 有

$$\frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \to N(0,1),$$

于是 $X \approx N(p, \dfrac{p(1-p)}{n})$. 利用枢轴变量法可以求出 $p$ 的置信度为 $1-\alpha$ 的置信区间为

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

这里的 $p$ 未知, 于是我们用 $\hat{p} = \bar{X}$ 代入上式, 最后得到 $p$ 的置信度为 $1-\alpha$ 的置信区间近似为

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

事实上, 注意到

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

只需将 $S$ 用 $S^*$ 代替就可以得到公式.

# 8 假设检验

前面我们简单介绍了参数估计. 从本章起, 我们将进入数理统计的另外一个重要的内容-假设检验. 简单地说, 假设检验就是对总体的某些未知特征提出假设, 并利用样本信息来推断该假设的正确性. 本章首先介绍假设检验的基本概念, 然后详细讨论正态总体的均值与方差的假设检验, 最后讲述拟合优度检验及独立性检验.

## 8.1 假设检验的基本概念

### 8.1.1 假设检验问题的提出

我们先看几个例子. **例 8.1** 某厂生产罐装食品, 每罐标准质量为 500 克. 根据经验认为质量服从正态分布, 标准差为 10. 现随机抽取 10 罐, 称得其质量 (单位: 克) 为

$$495, 510, 505, 498, 503, 492, 502, 512, 497, 506$$

试问该厂生产是否正常?

若用 $X$ 表示该厂生产的罐装食品的质量, 可以假定其服从正态分布 $N(\mu, 10^2)$. 于是该厂生产是否正常等于判断总体的均值 $\mu$ 是否等于 500. 我们将"$\mu = 500$"称为一个假设, 记为 $H_0 : \mu = 500$. 这在统计上称为**原假设**或**零假设**. 经检验如果拒绝 $H_0$, 则等价于接受"$\mu \neq 500$". 我们将其也看成一个假设, 记为 $H_1 : \mu \neq 50$, 称之为**备择假设**或**对立假设**. 我们常将它们写在一起. 如在例 8.1 中, 假设为

$$H_0 : \mu = 500; \quad H_1 : \mu \neq 500.$$

**例 8.2** 某厂生产一批产品, 按质量要求, 其次品率应低于 5%. 现从这批产品中抽取 50 件检验, 发现有 4 件次品. 试问该批次品能否出厂?

用 $p$ 表示这批产品的次品率, 现在的问题就是判断"$p \leq 0.05$"是否成立. 我们将其看成原假设, 记为 $H_0 : p \leq 0.05$. 此时对立假设记为 $H_1 : p > 0.05$. 合在一起写为

$$H_0 : p \leq 0.05; \quad H_1 : p > 0.05.$$

这类假设检验问题称为**单边检验** (原假设与对立假设各居一边), 而称假设 (8.1) 为双边检验 (对立假设分居原假设的两边). 又如果假设中只含有一个假设, 如假设 (8.1) 中的原假设, 则称为**简单假设**, 否则称为**复合假设**, 如假设 (8.2) 中的原假设.

如果检验的假设只涉及总体分布中的未知参数, 则称这种假设为**参数假设检验**, 除此之外的称为**非参数假设检验**. 我们看下面的例子.

**例 8.3** 某种建筑材料, 其抗断强度以往一直服从正态分布, 今改变配料方案, 希望确定其抗断强度的分布是否仍为正态分布?

上述例子就是一个非参数假设检验问题, 记 X 为建筑材料的抗断强度, 则原假设可以记为

$$H_0 : X\text{服从正态分布}.$$

### 8.1.2 假设检验的步骤

我们用例 8.1 来说明假设检验的具体操作过程.

首先还是先考虑均值 $\mu$ 的估计, 由第七章可知, 样本均值 $X$ 是一个合理的估计 (它既是矩估计, 又是极大似然估计). 从直观上, 若原假设 $H_0: \mu = 500$ 成立, 则样本均值 $X$ 与 $\mu_0 = 500$ 的差距不会太大. 这就要给出一个界限 $k$, 当 $|\bar{X} - \mu_0| \geq k$ 时, 则认为 $X$ 与 $\mu_0 = 500$ 的差距显著, 于是判断原假设 $H_0$ 不成立, 即对立假设 $H_1$ 成立; 当 $|\bar{X} - \mu_0| < k$, 则认为 $X$ 与 $\mu_0 = 500$ 的差距不显著, 于是判断原假设 $H_0$ 成立. 那么究竟应如何确定界限 $k$ 呢?

在原假设 $H_0$ 成立的情况下,$\bar{X} \sim N(\mu_0, \dfrac{\sigma^2}{n})$, 其中 $\sigma$ 已知 ($\sigma = 10$). 所以统计量

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

又 $|\bar{X} - \mu_0| \geq k$, 等价于 $|U| \geq \dfrac{k}{\sigma/\sqrt{n}}$. 于是若给定一个小概率 $\alpha$(通常取 $\alpha$ 为 0.1,0.05,0.01 等), 在 $H_0$ 成立的条件下, 由

$$P(|U| \geq \frac{k}{\sigma/\sqrt{n}}) = \alpha$$

即可确定界限 $k = \dfrac{\sigma}{\sqrt{n}} u_{\alpha/2}$.

例如, 取 $\alpha = 0.05$, 则 $u_{\alpha/2} = 1.96$. 又 $\sigma = 10, n = 10$ 得 $k = 6.2$. 由例 8.1 的数据得, 样本均值 $\bar{X} = 502$, 所以 $|\bar{X} - 500| = 2 < k = 6.2$, 故认为原假设 $H_0$ 成立.

我们将统计量 $U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ 称为**检验统计量**, $u_{\alpha/2} = \dfrac{k}{\sigma/\sqrt{n}}$ 称为**临界值**, $W = |U| \geq u_{\alpha/2}$ 称为**拒绝域**, 而称 $|U| < u_{\alpha/2}$ 为**接受域**.

通过以上分析, 将假设检验的基本步骤归纳如下:

(1) 根据问题提出原假设 $H_0$ 和对立假设 $H_1$;

(2) 构造一个合适的统计量 (往往由参数估计而来), 并在 $H_0$ 成立的条件下推导出该统计量的分布;

(3) 给出小概率 $\alpha$, 确定临界值和拒绝域 $W$;

(4) 由样本算出统计量的观察值, 若落在拒绝域 $W$, 则拒绝 $H_0$; 若落在接受域, 则接受 $H_0$.

### 8.1.3 假设检验的两类错误

从上面的讨论可以看出, 假设检验中可能犯以下两类错误: **第一类错误**是原假设 $H_0$ 正确, 但由于统计量的值落在拒绝域, 而拒绝了原假设 $H_0$, 这类错误也称为**弃真错误**; **第二类错误**是原假设 $H_0$ 不正确, 但统计量的值落在接受域, 而接受了原假设 $H_0$, 这类错误也称为**存伪错误**. 显然第一类错误的概率就是上面提到的小概率 $\alpha$, 即

$$P(\text{拒绝}H_0|H_0\text{ 真}) = P(U \in W|H_0\text{为真}) = \alpha,$$

有时也称 $\alpha$ 为**显著性水平**. 第二类错误我们一般记为 $\beta$, 即

$$P(\text{接受}H_0|H_1\text{为真}) = P(U \notin W|H_1\text{为真}) = \beta.$$

计算 $\beta$ 一般比较困难. 下面我们看一个例子, 初学可以略去.

**例 8.4** 设总体 $X$ 服从 $N(\mu,1)$, 现抽取样本 $X_1,X_2,\cdots,X_n$ 在给定显著性水平 $\alpha$ 下, 要检验假设

$$H_0:\mu=\mu_0;\quad H_1:\mu=\mu_1.$$

求第二类错误的概率 $\beta$.

**解** 由前一段可知, 检验统计量为 $U=\sqrt{n}(\bar{X}-\mu_0)$. 当原假设 $H_0$ 成立时,

$$U\sim N(0,1),$$

拒绝域 $W=|U|\geq u_{\alpha/2}$, 此时第一类错误为 $\alpha$; 当 $H_1$ 成立时, 不难看出

$$U\sim N(\sqrt{n}(\mu_1-\mu_0),1).$$

第二类错误的概率为

$$\begin{aligned}\beta&=P(\text{接受}H_0|H_1\text{为真})=P(U\notin W|H_1\text{为真})\\&=P(|U|<u_{\alpha/2}|\mu=\mu_1)\\&=\Phi[u_{\alpha/2}-\sqrt{n}(\mu_1-\mu_0)]-\Phi[-u_{\alpha/2}-\sqrt{n}(\mu_1-\mu_0)].\end{aligned}$$

在进行假设检验时, 当然希望犯两类错误的概率越小越好. 但在给定样本容量的情况下, 犯两类错误的概率不可能同时减小, 减少其中一个, 另外一个会增大. 基于这种情况,Neyman 与 Pearson(现代统计学的奠基人 K.Pearson 的儿子) 提出了一个原则: 在控制第一类错误 $\alpha$ 的前提下, 使犯第二类错误的概率 $\beta$ 尽量小. 这称为 Neyman-Pearson 原则. 因为在假设检验中, 原假设与备择假设的地位不是对等的, 往往会“保护”原假设. 也就是说, 只有在证据充分的情况下, 才拒绝原假设. 因此首先要控制第一类错误的概率, 然后才考虑第二类错误. 值得说明的是, 根据 Neyman-Pearson 的基本思想, 拒绝原假设是有充分证据的; 但接受原假设 $H_0$, 则未必有充分的理由. 只能说“目前还找不到拒绝 $H_0$ 的理由, 于是我们先接受 $H_0$”. 此时不能认为原假设 $H_0$ 一定正确.

### 8.1.4 *$p$ 值检验法

在假设检验中, 经常提到 $p$ 值. 本小节我们简单介绍一下 $p$ 值的概念. 这些内容初学可以略去不看.

对于假设检验

$$H_0:\theta\in\Theta_0;\quad H_1:\theta\in\Theta_1.$$

检验统计量为 $T=T(X_1,X_2,\cdots,X_n)$, 拒绝域 $W=|T|\geq\lambda$. 设 $x_1,x_2,\cdots,x_n$ 为样本观察值, 检验统计量 $T$ 的观察值为 $t=T(x_1,x_2,\cdots,x_n)$, 则 $p$ 值定义为

$$p(x_1,x_2,\cdots,x_n)=\sup_{\theta\in\Theta_0}P_\theta(|T|\geq|t|).$$

事实上,$p$ 值是基于样本观察值而计算出的拒绝原假设的概率 (当原假设为复合假设时, 取最大概率). 由于 $p$ 值依赖于样本观察值, 因此也是一个统计量. Fisher 曾称 $p$ 值为“反对原假设依据的强度”

. 当 $p$ 值小于显著性水平 $alpha$ 时, 我们拒绝原假设, 否则接受原假设. 可以这样形容 $p$ 值的作用, 它将假设检验的结果数量化了. 有了 $p$ 值, 我们不仅知道接受还是拒绝原假设, 而且还知道接受或拒绝原假设的“程度”. 随着计算机的普及,$p$ 值的计算变得十分方便, 很多统计软件的假设检验均给出了 $p$ 值.

## 8.2 正态总体均值的假设检验

### 8.2.1 单个正态总体 $N(\mu,\sigma^2)$ 均值 $\mu$ 的假设检验

设给定显著性水平 $\alpha$,$X_1,X_2,\cdots,X_n$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一组样本. 检验问题为

$$H_0:\mu=\mu_0;\quad H_1:\mu\neq\mu_0.$$

1.$\sigma^2$ 已知 ($u$ 检验)

这种情况我们在本章第一节已经做了详细的讨论, 检验统计量为 $U=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$. 在原假设 $H_0$ 成立的情况下,$U\sim N(0,1)$, 于是检验的拒绝域为

$$W=\left\{|U|=\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geq u_{\alpha/2}\right\}.$$

这种检验方法称为 $u$ 检验法. 如图所示.

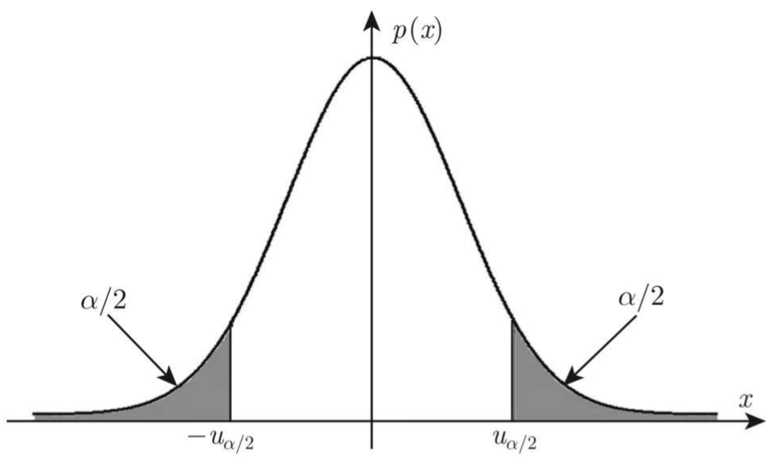


图 18

**例 8.6** 已知某零件的长度 (单位: 厘米) 服从 $N(\mu,1.1^2)$, 其标准长度 $\mu=32.05$. 现随机抽取 6 件, 测得长度为

$$32.56\quad 29.66\quad 31.64\quad 30.00\quad 31.87\quad 31.03$$

试问这批零件的长度是否符合产品的要求? 显著性水平 $\alpha=0.05$.

**解** 检验统计量为 $U=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$, 由样本观察值计算得 $\bar{X}=31.13$. 将 $\mu_0=32.05,\sigma=1.1$,$n=6$ 代入得检验统计量的观察值 $u=-2.05$. 又 $u_{0.025}=1.96$. 因此检验统计量的值落在拒绝域, 于是拒绝原假设 $H_0$. 即认为这批零件的长度不符合产品的要求.

下面我们顺便讲一下单边检验. 先看检验问题:

$$H_0: \mu = \mu_0; \quad H_1: \mu > \mu_0.$$

与问题 (8.6) 类似, 我们取检验统计量 $U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, 在原假设 $H_0$ 成立的情况下,$U$ 仍然服从 $N(0,1)$. 此时 $U$ 的值大, 则判断原假设 $H_0$ 不成立. 于是拒绝域 $W = \left\{ \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq u_\alpha \right\}$. 同理, 对于单边检验

$$H_0: \mu = \mu_0; \quad H_1: \mu < \mu_0,$$

拒绝域 $W = \left\{ \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -u_\alpha \right\}$. 现在讨论更一般的单边检验问题:

$$H_0: \mu \leq \mu_0; \quad H_1: \mu > \mu_0.$$

注意到问题 (8.10) 的原假设为复合假设. 我们仍取检验统计量 $U = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, 由于原假设是复合假设, 我们不能认为在原假设 $H_0$ 成立的情况下,$U \sim N(0,1)$. 但不难证明, 如果仍取拒绝域 $W = \left\{ \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq u_\alpha \right\}$, 此时第一类错误均小于显著性水平 (事实上, 当 $\mu = \mu_0$ 时, 第一类错误等于 $\alpha$, 而当 $\mu < \mu_0$ 时, 第一类错误小于 $\alpha$). 所以我们仍取拒绝域 $W = \left\{ \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq u_\alpha \right\}$. 对于检验

$$H_0: \mu \geq \mu_0; \quad H_1: \mu < \mu_0,$$

其拒绝域与问题 (8.9) 相同. 由此可见, 单边检验与双边检验只是拒绝域中不等式的方向不同, 其他并无太大差别, 因此在以后的检验问题中, 我们就不再对单边检验作详细介绍了.

2. $\sigma^2$ 未知 ($t$ 检验)

由于 $\sigma$ 未知, 我们用样本标准差 $S = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ 代替 $\sigma$. 于是检验统计量为 $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$. 由推论 6.1 可知, 在原假设 $H_0$ 成立的情况下,$T \sim t(n-1)$. 于是检验的拒绝域为 $W = \left\{ |T| = |\dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}| \geq t_{\alpha/2}(n-1) \right\}$. 上述检验方法利用 $t$ 统计量, 因此称为 $t$ 检验法. 如图所示.

**例 8.7** 已知某厂生产 10 欧姆的电阻. 假设电阻值 (单位: 欧姆) 服从正态分布, 方差未知. 现随机抽取 10 个电阻, 测得电阻值为

9.9　10.1　10.2　9.7　9.9　9.9　10　10.5　10.1　10.2

试问: 我们能否认为这批产品符合要求? 显著性水平 $\alpha = 0.05$.

**解**　首先计算得样本标准差 $S = 0.224$. 将 $\bar{X} = 10.05, \mu_0 = 10, n = 10$ 代入, 得检验统计量的观察值 $t = 0.707$. 又 $t_{0.025}(9) = 2.262$, 因此检验统计量的值落在接受域, 于是接受原假设 $H_0$. 即认为这批产品符合要求.

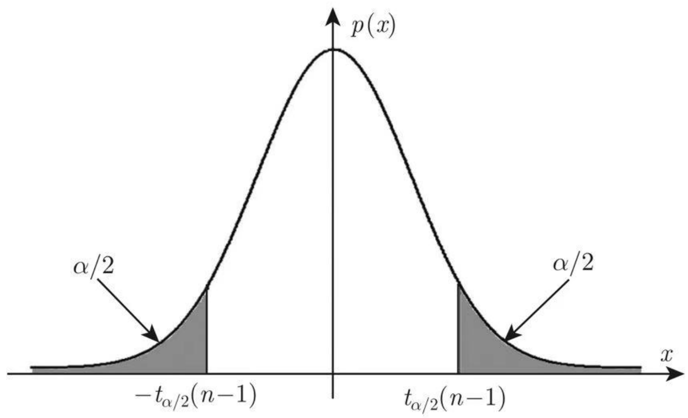


图 19

### 8.2.2 两个正态总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ 的均值差的检验

设 $X_1, X_2, \cdots, X_{n_1}$ 是来自正态总体 $N(\mu_1, \sigma_1^2)$ 的一个样本,$Y_1, Y_2, \cdots, Y_{n_2}$ 为来自正态总体 $N(\mu_2, \sigma_2^2)$ 的一个样本, 且两样本相互独立. 记 $X$ 和 $Y$ 分别为它们的样本均值,$S_1^2$ 和 $S_2^2$ 分别为它们的样本方差. 检验问题为

$$H_0: \mu_1 = \mu_2; \quad H_1: \mu_1 \neq \mu_2.$$

1.$\sigma_1^2$,$\sigma_2^2$ 均已知

取检验统计量

$$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

当原假设 $H_0$ 成立的情况下,$U \sim N(0,1)$. 此时统计量 $U$ 的绝对值不会太大. 因此, 在给定显著性水平 $\alpha$ 下, 拒绝域

$$W = \left\{ |U| \geq u_{\alpha/2} \right\}.$$

**例 8.8** 设甲、乙两台机床生产同一产品, 今从甲机床生产的产品中抽取 30 件产品, 测得平均质量为 130 克. 从乙机床生产的产品中抽取 40 件产品, 测得平均质量为 125 克. 假定两台机床生产的产品质量均服从正态分布, 方差为 $\sigma_1^2 = 60, \sigma_2^2 = 80$. 问在显著性水平 $\alpha = 0.05$ 下, 两台机床生产的产品质量是否有显著差异?

**解** 将 $\bar{X} = 130, \bar{Y} = 125, \sigma_1^2 = 60, \sigma_2^2 = 80, n_1 = 30, n_2 = 40$ 代入, 得检验统计量的观察值 $u = -2.5$. 又 $u_{0.025} = 1.96$. 因此检验统计量的值落在拒绝域, 于是拒绝原假设. 即认为两台机床生产的产品质量有显著差异.

2.$\sigma_1^2$,$\sigma_2^2$ 均未知但相等

由推论 6.3, 取检验统计量

$$T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{\bar{X} - \bar{Y}}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}}.$$

当原假设 $H_0$ 成立的情况下,$T \sim t(n_1+n_2-2)$. 此时统计量的值应在 0 的附近. 因此, 在给定显著性水平 $\alpha$ 下, 拒绝域

$$W = \{|T| \geq t_{\alpha/2}(n_1+n_2-2)\}.$$

**例 8.9** 设某工厂利用两条自动化流水线灌装矿泉水. 随机地抽取样本 $X_1, X_2, \cdots, X_{12}$ 和 $Y_1, Y_2, \cdots, Y_{17}$, 它们是每瓶矿泉水的体积 (单位: 毫升). 计算得样本均值 $\bar{X} = 501.1$ 和 $\bar{Y} = 499.7$, 样本方差 $S_1^2 = 2.4, S_2^2 = 4.7$. 假定两条流水线生产的矿泉水体积均服从正态分布, 且它们的方差相等. 问在显著性水平 $\alpha = 0.05$ 下, 两条生产线生产的产品是否有显著差异?

**解** 该题与例 7.18 相同, 只是在例 7.18 中是求均值差的置信度为 95% 的置信区间. 将样本观察值代入得 $t = 1.914$, 又 $t_{0.025}(27) = 2.05$. 因此检验统计量的值落在接受域, 于是接受原假设 $H_0$, 即认为两条生产线生产的产品没有显著差异. 该结论与例 7.18 相同.

### 8.2.3 基于成对数据的假设检验

上面讨论的两个正态总体均值的比较检验中, 这两个正态总体是相互独立的. 但在实际中, 并非总是如此. 可能这两个正态总体的样本来自同一个总体的重复测量, 它们是成对出现且是相关的. 例如, 考虑两种药品的效果, 测试了 $n$ 个病人服药后的状况, 记为 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$. 其中 $(X_i, Y_i)$ 表示第 $i$ 个病人服用两种药后的数据, 它们是有关系的. 这样的数据就称为成对数据. 现令 $Z_i = X_i - Y_i, i = 1, 2, \cdots, n$. 假定 $Z_i \sim N(\mu, \sigma^2)$, 且相互独立. 下面检验

$$H_0: \mu = 0; \quad H_1: \mu \neq 0,$$

这是一个正态总体均值的检验. 由于方差未知, 利用 $t$ 检验法, 在显著性水平 $\alpha$ 下, 拒绝域 $W = \left\{\left|\dfrac{\bar{Z}}{S_Z/\sqrt{n}}\right| \geq t_{\alpha/2}(n-1)\right\}$. 其中 $\bar{Z} = \dfrac{1}{n}\sum_{i=1}^{n} Z_i, S_Z^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$. 这个检验通常称为**成对 $t$ 检验**.

**例 8.10** 比较两种安眠药 A,B 的疗效, 以 10 个失眠患者作为试验对象. 对每个患者服药后, 记录其延长的睡眠时间 (单位: 小时), 数据如下:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | 1.9 | 0.8 | 1.1 | 0.1 | -0.1 | 4.4 | 5.5 | 1.6 | 4.6 | 3.4 |
| $B$ | 0.7 | -1.6 | -0.2 | -1.2 | -0.1 | 3.4 | 3.7 | 0.8 | 0 | 2.0 |

问在显著性水平 $\alpha = 0.01$ 下, 两种药品是否有显著差异?

**解** 令 $Z_i = X_i - Y_i, i = 1, 2, \cdots, 10$. 数据如下:

1.2  2.4  1.3  1.3  0  1.0  1.8  0.8  4.6  1.4

利用 $t$ 检验法, 计算得 $\bar{Z} = 1.58, S_Z = 1.23$. 代入得统计量的观察值 $t = \dfrac{\bar{Z}}{S_Z/\sqrt{n}} = 4.06$. 又 $t_{0.005}(9) = 3.25$, 现 $|t| = 4.06 > t_{0.005}(9) = 3.25$. 故拒绝原假设, 即认为两种药品有显著差异.

## 8.3 正态总体方差的假设检验

### 8.3.1 单个正态总体 $N(\mu, \sigma^2)$ 方差 $\sigma^2$ 的假设检验

设给定显著性水平 $\alpha, X_1, X_2, \cdots, X_n$ 为来自总体 $X \sim N(\mu, \sigma^2)$ 的一组样本. 检验问题为

$$H_0: \sigma^2 = \sigma_0^2; \quad H_1: \sigma^2 \neq \sigma_0^2.$$

这里假定 $\mu$ 未知. 由定理 6.2 知, 取检验统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}.$$

在原假设 $H_0$ 成立的情况下,$\chi^2 \sim \chi^2(n-1)$. 此时 $\chi^2$ 的值不应太大或太小. 于是, 在给定显著性水平 $\alpha$ 下, 拒绝域

$$W = \{\chi^2 \leq \chi^2_{1-\alpha/2}(n-1)\} \cup \{\chi^2 \geq \chi^2_{\alpha/2}(n-1)\}.$$

上述检验称为 $\chi^2$ 检验法. 如图所示.

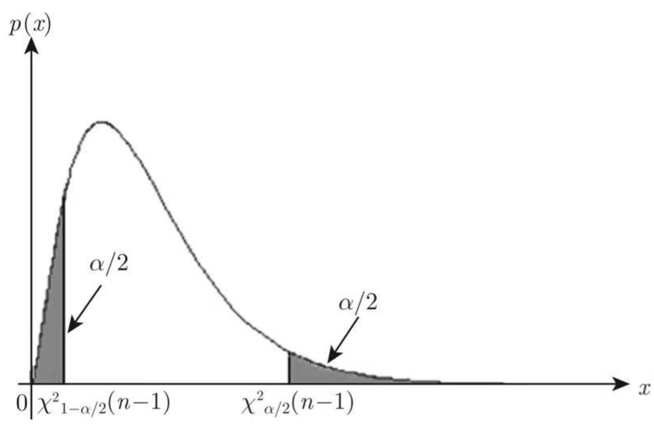


图 20

**例 8.11** 设某厂生产的铜丝的折断力服从正态分布. 现从中随机抽取 10 根进行检验, 测得其折断力 (单位: 牛顿) 为

575 576 570 569 572 582 577 580 572 585

试问在显著性水平 $\alpha = 0.05$ 下, 能否认为该厂生产的铜丝的折断力的方差为 64?

**解** 由样本计算得 $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2} = \dfrac{251.6}{64} = 3.93$. 又 $\chi^2_{0.975}(9) = 2.70, \chi^2_{0.025}(9) = 19.02$. 检验统计量的值落在接受域. 故接受原假设, 认为该厂生产的铜丝的折断力的方差为 64.

### 8.3.2 两个正态总体 $N(\mu_1,\sigma_1^2), N(\mu_2,\sigma_2^2)$ 的方差比 $\sigma_1^2/\sigma_2^2$ 的假设检验

设 $X_1, X_2, \cdots, X_{n_1}$ 是来自正态总体 $N(\mu_1,\sigma_1^2)$ 的一个样本,$Y_1, Y_2, \cdots, Y_{n_2}$ 为来自正态总体 $N(\mu_2,\sigma_2^2)$ 的一个样本, 且两样本相互独立. 检验问题为

$$H_0: \sigma_1^2 = \sigma_2^2; \quad H_1: \sigma_1^2 \neq \sigma_2^2,$$

同样我们假定均值 $\mu_1, \mu_2$ 未知. 由推论 6.2, 统计量

$$F = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \sim F(n_1-1, n_2-1),$$

在原假设 $H_0$ 成立的情况下, 有 $F = \dfrac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$. 在给定显著性水平 $\alpha$ 下, 拒绝域

$$W = \{F \leq F_{1-\alpha/2}(n_1-1, n_2-1)\} \cup \{F \geq F_{\alpha/2}(n_1-1, n_2-1)\}$$

上述检验称为 $F$ 检验法. 如图所示.

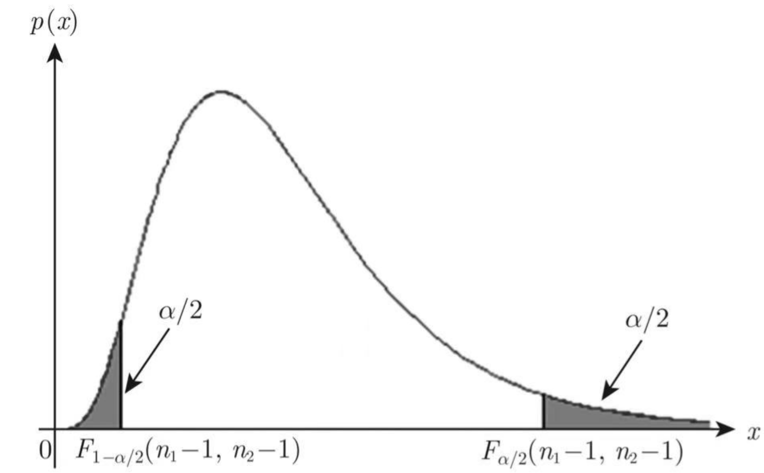


图 21

**例 8.12** 设甲、乙两台机床加工同一种轴, 现随机抽取若干根, 测得直径 (单位: 毫米) 为

甲 : 20.519.719.820.420.120.019.019.9

乙 : 19.720.820.519.819.420.619.2

假定甲乙机床加工轴的直径均服从正态分布. 试问甲、乙两台机床加工的精度有无显著差异?

**解** 由样本计算得 $S_1^2 = 0.216, S_2^2 = 0.397$. 所以 $F = \dfrac{S_1^2}{S_2^2} = \dfrac{0.216}{0.397} = 0.544$. $F_{0.975}(7,6) = \dfrac{1}{F_{0.025}(6,7)} = \dfrac{1}{5.12} = 0.195, F_{0.025}(7,6) = 5.7$. 因为 $F_{0.975}(7,6) = 0.195 < F = 0.544 < F_{0.025}(7,6) = 5.7$, 所以接受原假设 $H_0$, 即认为甲、乙两台机床加工的精度无显著差异.

## 8.4 * 拟合优度检验

前面两节讨论的假设检验都是在假定总体分布为正态分布的前提下进行的检验, 这称为参数假设检验. 在实际问题中, 我们常常不知道总体的分布类型, 这就需要对总体的分布类型进行推断, 如例 8.3 中检验分布是否为正态分布. 这种检验称为**分布拟合优度检验**. 一般地, 假设如下:

$$H_0: F(x) = F_0(x;\theta),$$

其中 $F_0$ 为某个已知的分布函数,$\theta = (\theta_1, \theta_2, \cdots, \theta_r)'$ 为未知参数.

下面介绍皮尔逊 $\chi^2$ 拟合优度检验. 其基本思想就是利用事件的频率与概率之间的偏差构造检验统计量. 我们先看最简单的问题:

$$H_0: P(X = x_i) = p_i, i = 1, 2, \cdots, k,$$

其中 $p_i, i = 1, 2, \cdots, k$ 为已知的概率, 满足 $\sum\limits_{i=1}^{k} p_i = 1$ 设 $X_1, X_2, \cdots, X_n$ 是一组样本. 构造皮尔逊 $\chi^2$ 统计量的步骤如下:

(1) 计算 $(X_1, X_2, \cdots, X_n)$ 中取 $x_i$ 的实际频数 $n_i, i = 1, 2, \cdots, k$, 即

$$n_i = \{X_1, X_2, \cdots, X_n \text{中取} x_i \text{的个数};$$

(2) 计算实际频数与理论频数的偏差平方和

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i},$$

这称为皮尔逊 $\chi^2$ 统计量, 它度量了实际数据与理论分布 $H_0$ 的拟合程度.$\chi^2$ 统计量的值越大, 则拟合程度越差. 可以证明, 当假设 (8.18) 为真时,$\chi^2$ 统计量的渐进分布为自由度 $k-1$ 的 $\chi^2$ 分布.

(3) 在给定显著性水平 $\alpha$ 下, 拒绝域为

$$W = \{\chi^2 \geq \chi^2_\alpha(k-1)\}.$$

下面我们考虑一般的假设检验问题 (8.17). 具体步骤如下:

(1) 将样本空间分为 $k$ 个互不相交的事件 $A_1, A_2, \cdots, A_k$.

(2) 计算样本 $(X_1, X_2, \cdots, X_n)$ 落在事件 $A_i$ 的实际频数 $n_i, i = 1, 2, \cdots, k$,

$$n_i = X_1, X_2, \cdots, X_n \text{落在事件} A_i \text{的个数}.$$

(3) 计算每个事件 $A_i$ 上的理论频数. 若参数 $\theta$ 未知, 先算出 $\theta$ 的极大似然估计, 然后计算理论上样本落在事件 $A_i$ 的概率 $\hat{p}_i = P(X \in A_i | \theta = \hat{\theta}), i = 1, 2, \cdots, k$. 最后得到每个事件上的理论频数为 $n\hat{p}_i$. 也就是说, 应该有 $n\hat{p}_i$ 个样本落在事件 $A_i$ 中.

(4) 计算实际频数与理论频数的偏差平方和

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

(5) 在给定显著性水平 $\alpha$ 下, 下面定理给出了检验的拒绝域

$$W = \{\chi^2 \geq \chi^2_\alpha(k-r-1)\}.$$

**定理 8.1(Pearson-Fisher 定理)** 当原假设 $H_0$ 为真时, 公式 (8.21) 定义的统计量的极限分布为自由度为 $k-r-1$ 的 $\chi^2$ 分布.

值得注意的是, 皮尔逊拟合优度检验只适合大样本情形, 通常要求 $n \geq 50$. 将样本空间划分为事件, 要求每个事件上的理论频数不应太小, 一般每个 $n\hat{p}_i$ 最好不小于 5, 否则需将相邻的事件合并. 同时 $k$ 不应太大或太小, 一般在 $4 \sim 20$.

## 8.5 * 独立性检验

在实际应用中我们常会遇到这样的问题: 判断两个随机变量之间是否独立. 例如, 患肺癌与吸烟是否有关, 患高血压与冠心病是否相关等. 这类问题我们可以用本章第四节的 $\chi^2$ 拟合优度加以检验.

设 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ 为二维总体 $(X, Y)$ 的一个样本, 我们的问题是检验:

$$H_0 : X\ Y\ \ .$$

分别将随机变量 $X$ 与 $Y$ 的取值范围分成 $r$ 个与 $s$ 个互不相交的区间 $A_1, A_2, \cdots, A_r$ 和 $B_1, B_2, \cdots, B_s$. 用 $n_{ij}$ 表示指标 $X$ 落在 $A_i$, 指标 $Y$ 落在 $B_j$ 的样本点的个数. 又记 $n_{i.} = \sum_{j=1}^{s} n_{ij}, i-1, 2, \cdots, r; n_{.j} = \sum_{i=1}^{r} n_{ij}, j = 1, 2, \cdots, s$. 我们将这些数据用表 10 给出, 此表称为列联表.

| | $B_1$ | $B_2$ | $\cdots$ | $B_j$ | $\cdots$ | $B_s$ | 行和 |
|---|---|---|---|---|---|---|---|
| $A_1$ | $n_{11}$ | $n_{12}$ | $\cdots$ | $n_{1j}$ | $\cdots$ | $n_{1s}$ | $n_{1.}$ |
| $A_2$ | $n_{21}$ | $n_{22}$ | $\cdots$ | $n_{2j}$ | $\cdots$ | $n_{2s}$ | $n_{2.}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | | $\cdots$ | | $\cdots$ | $\cdots$ |
| $A_i$ | $n_{i1}$ | $n_{i2}$ | $\cdots$ | $n_{ij}$ | $\cdots$ | $n_{is}$ | $n_{i.}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | | $\cdots$ | | $\cdots$ | $\cdots$ |
| $A_r$ | $n_{r1}$ | $n_{r2}$ | $\cdots$ | $n_{rj}$ | $\cdots$ | $n_{rs}$ | $n_{r.}$ |
| 列和 | $n_{.1}$ | $n_{.2}$ | $\cdots$ | $n_{.j}$ | $\cdots$ | $n_{.s}$ | $n$ |

表 10: $r \times s$ 列联表

记 $p_{ij} = P(X \in A_i, Y \in B_j), i = 1, \cdots, r; j = 1, \cdots, s; p_{i.} = P(X \in A_i), i = 1, \cdots, r; p_{.j} = P(Y \in B_i), j = 1, \cdots, s$.

于是独立性检验就相当于检验:

$$H_0 : p_{ij} = p_{i\cdot} p_{\cdot j}, i = 1, \cdots, r; j = 1, \cdots, s.$$

该假设中含有 $r+s$ 个未知参数 $p_{i\cdot j} i = 1, \cdots, r; p_{\cdot j}, j = 1, \cdots, s$. 但显然有 $\sum_{i=1}^{r} p_{i.} = 1, \sum_{j=1}^{s} p_{.j} = 1$, 因此仅有 $r+s-2$ 个独立的未知参数. 下面用 $\chi^2$ 检验法对其加以检验. 首先用极大似然估计对未知参数进行估计. 当原假设成立时, 似然函数为

$$\begin{aligned} L &= \prod_{i=1}^{r} \prod_{j=1}^{s} p_{ij}^{n_{ij}} = \prod_{i=1}^{r} \prod_{j=1}^{s} (p_{i.} p_{.j})^{n_{ij}} = \prod_{i=1}^{r} p_{i.}^{n_{i.}} \prod_{j=1}^{s} p_{.j}^{n_{.j}} \\ &= \prod_{i=1}^{r-1} p_{i.}^{n_{i.}} \left(1 - \sum_{i=1}^{r-1} p_{i.}\right)^{n_{r.}} \prod_{j=1}^{s-1} p_{.j}^{n_{.j}} \left(1 - \sum_{j=1}^{s-1} p_{.j}\right)^{n_{.s}} \end{aligned}$$

对数似然为

$$\ln L = \sum_{i=1}^{r-1} n_{i.} \ln p_{i.} + n_{r.} \ln \left(1 - \sum_{i=1}^{r-1} p_{i.}\right) + \sum_{j=1}^{s-1} n_{.j} \ln p_{.j} + n_{.s} ln(1 - \sum_{j=1}^{s-1} p_{.j})$$

解似然方程组

$$\begin{aligned} \frac{\partial \ln L}{\partial p_{i\cdot}} &= \frac{-n_r}{1 - \sum_{i=1}^{r-1} p_{i\cdot}} + \frac{n_{i\cdot}}{p_{i\cdot}} = 0, i = 1, \cdots, r-1; \\ \frac{\partial \ln L}{\partial p_{\cdot j}} &= \frac{-n.g}{1 - \sum_{j=1}^{s-1} p_{\cdot j}} + \frac{n_{\cdot j}}{p_{\cdot j}} = 0, j = 1, \cdots, s-1. \end{aligned}$$

得

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}, i = 1, \cdots, r; \hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}, j = 1, \cdots, s.$$

将其代入 $\chi^2$ 检验统计量

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n\hat{p}_{i.}\hat{p}_{.j})^2}{n\hat{p}_{i.}\hat{p}_{.j}} = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n}$$

当原假设成立时, 它的渐进分布为自由度 $rs - (r+s-2) - 1 = (r-1)(s-1)$ 的 $\chi^2$ 分布. 于是, 在给定显著性水平 $\alpha$ 下, 检验的拒绝域为 $W = \{\chi^2 \geq \chi^2_\alpha(r-1)(s-1))\}$.