



天文数据处理方法

万晓生

xswan@nju.edu.cn



- 概率统计基础
- 误差概论和最小二乘法
- 回归分析
- 谱分析基础
- 平滑与滤波
- 随机信号的功率谱估计（略）



观测是获得天文资料的主要手段

70年代

紫外

IUE

80年代

红外

IRAS 红外天文卫星

HEAO 高能天文台序列



地基巡天数据

DPOSS
USNO-A2
SDSS
2MASS
NVSS
FIRST

地基天文台观测数据

NOAO
NRAO
ESO
NAOC

空间天文台观测数据

MAST
IRSA
HEASARC



IRAF: Image Reduction and Analysis Facility

IDL: Interactive Data Language

AIPS: Astronomical Image Processing System

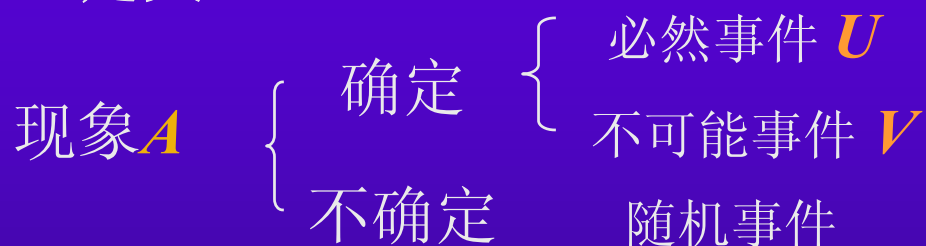
MIDAS: Munich Image Data Analysis System

第一章 概率统计基础

§ 随机事件及其概率

一 随机事件

1. 定义



随机试验存在不同的可能结果：

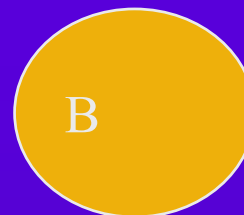
- 基本事件（不可分）
- 复合事件（基本事件的组合）
- 基本事件空间 Ω

2. 关系

1) 事件的包含

A 发生 $\longrightarrow$ B 发生

$$B \supset A$$



2) 事件的相等

$$A \supset B \quad B \supset A \longrightarrow A = B$$

3) 互不相容(互斥)

A 、 B 不同时发生

$$A \cap B = \emptyset$$



4) 对立事件

$$\begin{cases} A \cap B = \emptyset \\ A \cup B = U \end{cases}$$



5) 事件的完备群

$A_1, A_2, \dots, A_n$ 互不相容, 能且只能出现一个



3. 运算

1) 事件之和

$$A + B$$

$$A \cup B$$

$$A \cup B$$

2) 事件之积

$$AB$$

$$A \cap B$$

$$A \cap B$$

3) 事件之差

$$A - B$$

$$A \bar{B}$$

$$A - B$$

4) 对偶律

$$\overline{\bigcup A_i} = \bigcap \bar{A}_i$$

二 概率

$P(A)$: 事件 A 在一定条件下出现的概率

1. 古典概率

n 种结果 n_A 次有利于 A 发生

$$P(A) = n_A/n$$

2. 统计概率

n 次试验 m 次有利于 A 发生

$$W(A) = m/n \quad \text{频率}$$

$$P(A) = \lim W(A)$$

三 概率运算

1. 概率加法

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

2. 概率减法

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

3. 概率乘法

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

条件概率

概率统计基础

例：箱内有3个白球， 2个红球， 甲乙两人依次从中取一球， 甲先取， 若已知甲取得(a)白球， (b)红球， 试在甲取后放还及不放还两种情况下， 球乙取得红球的概率。

条件频率 $W(A | B) = n_{AB} / n_B$

条件概率 $P(A | B) = \lim W(A | B) = P(AB) / P(B)$

若事件 A 的发生不受事件 B 是否发生的影响， 即 $P(A|B) = P(A)$ ， 则称 A 、 B 相互独立。

独立事件 A 、 B $P(AB) = P(A) P(B)$

$$\begin{aligned} P(A_1 + A_2 + A_3 + \cdots + A_n) &= 1 - P(\overline{A_1 + A_2 + A_3 + \cdots + A_n}) \\ &= 1 - P(\bigcap_{k=1}^n \overline{A_k}) \\ &= 1 - \bigcap_{k=1}^n P(\overline{A_k}) \end{aligned}$$

概率统计基础

例：市场上的灯泡，甲厂有70%的占有率，乙厂有30%的占有率。而甲厂灯泡的合格率为95%，乙厂灯泡的合格率为80%。问市场上灯泡的合格率是多少？合格灯泡是甲厂产品的概率是多少？

解：市场合格率 $P(A)$ 甲厂 B_1 乙厂 B_2

$$P(B_1) = 0.7 \qquad P(A | B_1) = 0.95$$

$$P(B_2) = 0.3 \qquad P(A | B_2) = 0.80$$

$$A = AB_1 + AB_2$$

$$P(A) = P(AB_1) + P(AB_2)$$

$$= P(B_1) P(A | B_1) + P(B_2) P(A | B_2)$$

$$= 0.905$$

$$P(B_1 | A) = P(B_1) P(A | B_1) / P(A) = 0.736$$

概率统计基础

全概率公式 若事件 A 能且只能与互不相容事件完备群 $B_1, B_2, \dots, B_n$ 之一同时发生, 则有

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

贝叶斯(Bayes)公式 若事件 A 能且只能与互不相容事件完备群 $B_1, B_2, \dots, B_n$ 之一同时发生, 则在 A 发生的条件下, 事件 B_i 发生的概率为

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

§ 随机变量及其分布

一 随机变量**RV**: 表示随机试验结果的量 $\xi \eta \zeta$

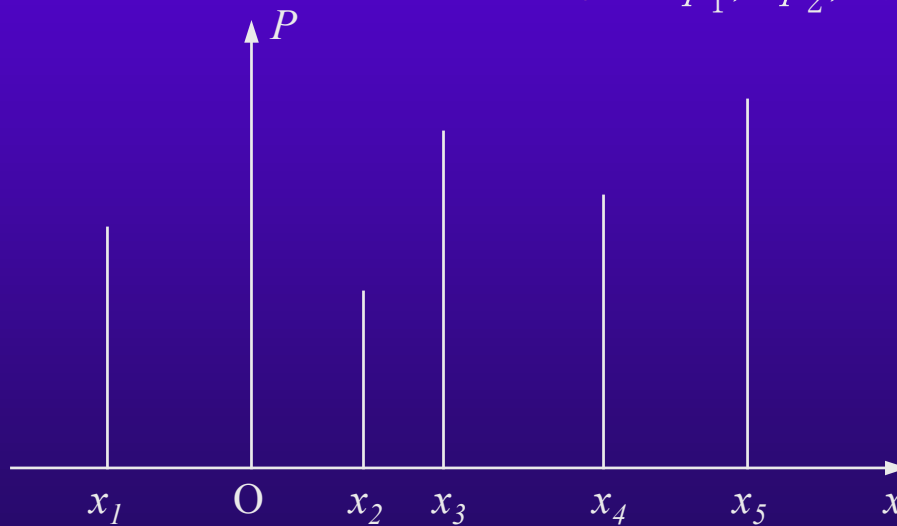
1. 离散分布与概率分布律

随机变量 ξ

有限个、可列的数值

概率分布律

$$\left\{ \begin{array}{ll} \xi: & x_1, x_2, \dots \\ P: & p_1, p_2, \dots \end{array} \right.$$



$$\left\{ \begin{array}{l} p_i \geq 0 \quad i = 1, 2, \dots \\ \sum p_i = 1 \end{array} \right.$$

2. 连续型随机变量与分布函数

连续型随机变量：若 ξ 是一随机变量， x 是任意实数，则定义 ξ 取值小于或等于 x 这样一个随机事件的概率为随机变量 ξ 的分布函数 $F(x)$ ，即

$$F(x) = P(\xi \leq x)$$

性质：① $0 \leq F(x) \leq 1$ $F(-\infty) = 0$ $F(\infty) = 1$

② $F(x)$ 为非降函数， $x_1 < x_2$ ，有 $F(x_1) < F(x_2)$

$$P(\xi \leq x_2) = P(\xi \leq x_1) + P(x_1 < \xi \leq x_2)$$

$$P(x_1 < \xi \leq x_2) = P(\xi \leq x_2) - P(\xi \leq x_1)$$

$$= F(x_2) - F(x_1) > 0$$

③ $F(x)$ 为右连续 $F(x+0) = F(x)$



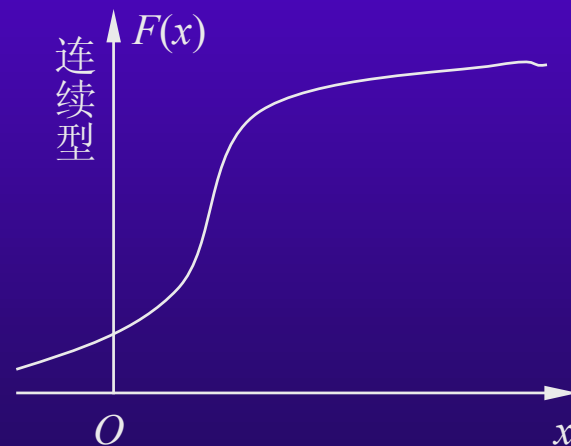
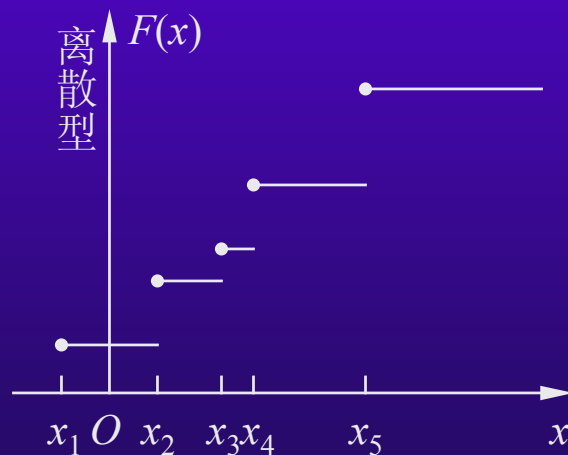
3. 离散型随机变量的分布函数_i

$$F(x) = P(\xi < x) = \sum_{x_i < x} P(\xi = x_i) = \sum_{x_i < x} P_i$$

$$F(x) = \begin{cases} 0 & x < x_1 \\ p_1 & x_1 \leq x < x_2 \\ p_1 + p_2 & x_2 \leq x < x_3 \\ \dots \dots & \dots \dots \end{cases}$$

性质：① $F(x)$ 为一右连续的阶梯函数

② $F(x)$ 有跃度，即 $P(\xi = x_i)$



分布函数

例：离散型随机变量 ξ 的分布律为

ξ	0	1	2
P	1/3	1/6	1/2

求：(1) ξ 的分布函数；

(2) $P(\xi \leq 1/2)$, $P(1 < \xi < 3/2)$ 。

解：

$$(1) \quad F(x) = P(\xi \leq x) = \begin{cases} 0 & x < 0 \\ 1/3 & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

$$(2) \quad P(\xi \leq 1/2) = F(1/2) = 1/3$$

$$P(1 < \xi < 3/2) = F(3/2) - F(1) = 1/2 - 1/2 = 0$$

二 分布密度函数

若连续随机变量 ξ 的分布函数 $F(x)$ 连续可微, 则定义

$$f(x) = \mathrm{d}F(x)/\mathrm{d}x$$

为 ξ 的分布密度函数或概率密度函数。

性质: ① $f(x) \geq 0$;

$$\textcircled{2} P(x_1 < \xi < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) \mathrm{d}x$$

$$\textcircled{3} \int_{-\infty}^{\infty} f(x) \mathrm{d}x = 1$$

例:

$$f(x) = \begin{cases} k e^{-3x} & x > 0 \\ 0 & x \leq 0 \end{cases}, \text{ 求 } P(\xi > 0.1)$$

解:

$$\textcircled{1} \text{ 求 } k \quad \int_{-\infty}^{\infty} f(x) \mathrm{d}x = \int_0^{\infty} k e^{-3x} \mathrm{d}x = 1 \quad k = 3$$

$$\begin{aligned} \textcircled{2} P(\xi > 0.1) &= 1 - P(\xi \leq 0.1) \\ &= 1 - \int_{-\infty}^{0.1} f(x) \mathrm{d}x \\ &= 0.74 \end{aligned}$$

三 随机变量函数的分布

设随机变量 ξ 的概率密度为 $f(x)$, $\eta=\varphi(\xi)$ 是 ξ 的单调连续函数, 且处处可导, 则 η 的概率密度函数 $g(y)$ 为

$$g(y) = f[\psi(y)]|\psi'(y)|$$

其中 $\psi(y)$ 是 $\varphi(x)$ 的反函数, $\psi'(y)$ 是 $\psi(y)$ 的导数。

证明: (1) 若 $\varphi(x)$ 是单调增加, 则 $\xi < x \Leftrightarrow \eta < y$

$$G(y) = P(\eta < y) = P(\xi < x) = \int_{-\infty}^x f(x) dx$$

$$g(y) = G'(y) = f[\Psi(y)]\Psi'(y) \quad \Psi'(y) > 0$$

(2) 若 $\varphi(x)$ 是单调减少, 则 $\xi > x \Leftrightarrow \eta < y$

$$G(y) = P(\eta < y) = P(\xi > x) = 1 - \int_{-\infty}^x f(x) dx$$

$$g(y) = G'(y) = -f[\Psi(y)]\Psi'(y) \quad \Psi'(y) < 0$$

由(1)和(2), 有

$$g(y) = f[\psi(y)]|\psi'(y)|$$

概率统计基础

例：由统计物理学知道分子运动速度的绝对值 X 服从麦克斯韦分布，其概率密度为

$$f(x) = \begin{cases} \frac{4x^2}{a^3\sqrt{\pi}} e^{-\frac{x^2}{a^2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

其中 $a > 0$ 为常数，求分子动能 $Y = \frac{1}{2}mX^2$ (m 为分子的质量) 的概率密度。

解：

$$\varphi(x): \quad y = \frac{1}{2}mx^2$$

$$\psi(y): \quad x = \sqrt{2y/m}$$

$$\psi'(y): \quad x' = \sqrt{1/2ym}$$

$$g(y) = f[\psi(y)]|\psi'(y)| = \begin{cases} \frac{4\sqrt{2y}}{m^{3/2}a^3\sqrt{\pi}} e^{-\frac{2y}{a^2m}} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

§ 多维随机变量及其分布

一 二维分布

1. 离散型随机变量

(ξ, η) 有限或可列

$$P(\xi=x_i, \eta=y_j) = P_{ij} \quad i, j = 1, 2, 3, \dots$$

联合概率分布律:

$$(1) \quad P_{ij} \geq 0 \quad \text{非负性;}$$

$$(2) \quad \sum_j P_{ij} = 1 \quad \text{归一化。}$$

边缘分布律: ξ, η

$$(1) \quad p_i = P(\xi = x_i) = \sum_j P_{ij} \quad i = 1, 2, 3, \dots$$

$$(2) \quad p_j = P(\eta = y_j) = \sum_i P_{ij} \quad j = 1, 2, 3, \dots$$

$$\xi \text{ 的条件概率} \quad P(\xi = x_i | \eta = y_j) = \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} = \frac{P_{ij}}{P_{\bullet j}}$$

2. 连续型随机变量

分布函数 $F(x,y) = P(\xi < x, \eta < y)$

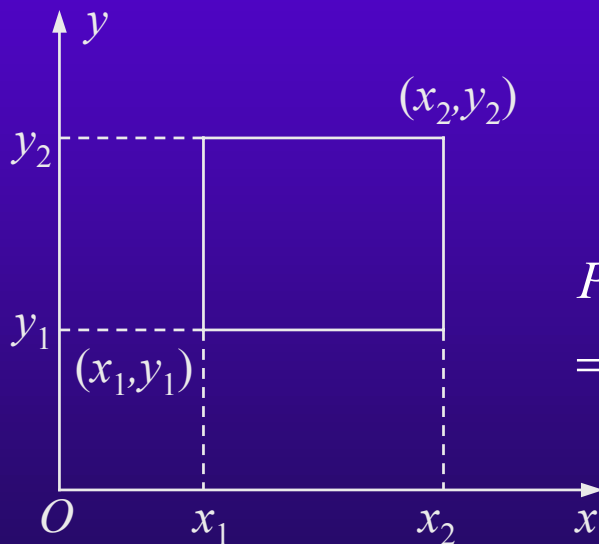
性质: ① $x_2 > x_1 \Rightarrow F(x_2, y) > F(x_1, y),$
 $y_2 > y_1 \Rightarrow F(x, y_2) > F(x, y_1);$

② $0 \leq F(x, y) \leq 1$ $F(-\infty, y) = 0$

$F(x, -\infty) = 0$

$F(-\infty, -\infty) = 0$

$F(+\infty, +\infty) = 1$



$$P(x_1 < \xi < x_2, y_1 < \eta < y_2) \\ = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

概率统计基础

分布密度函数 $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$

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

性质: ① $\forall x, y \quad f(x, y) \geq 0;$

② 归一化 $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1;$

③ $P\{(\xi, \eta) \in \mathcal{D}\} = \iint_{\mathcal{D}} f(x, y) dx dy$

边缘分布函数 $F_{\xi}(x) \stackrel{\text{def}}{=} P(\xi < x, -\infty < \eta < \infty)$

$$\begin{aligned} &= F(x, \infty) \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f(x, y) dx dy \\ &= \int_{-\infty}^x f_{\xi}(x) dx \end{aligned}$$

边缘分布密度

同样

$$f_{\xi}(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$F_{\eta}(y) = \int_{-\infty}^y f_{\eta}(y) dy$$

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

概率统计基础

相互独立

$$F(x, y) = F_{\xi}(x) \cdot F_{\eta}(y)$$

$$f(x, y) = f_{\xi}(x) \cdot f_{\eta}(y)$$

条件分布

$$f_{\xi|\eta}(x|y) = f(x, y) / f_{\eta}(y)$$

$$f_{\eta|\xi}(y|x) = f(x, y) / f_{\xi}(x)$$

$$P_{\xi|\eta}(x|y) = \lim_{\varepsilon \rightarrow 0} P(\xi < x | y - \varepsilon < \eta < y + \varepsilon)$$

$$F(x|y) = F_{\xi}(x)$$

$$F(y|x) = F_{\eta}(y)$$

$$f_{\xi|\eta}(x|y) = f_{\xi}(x)$$

$$f_{\eta|\xi}(y|x) = f_{\eta}(y)$$

例： 已知二维随机变量 (ξ, η) 的概率密度为 $f(x, y)$ ，求 $\zeta = \xi + \eta$ 的分布密度。

解：

$$F_{\zeta}(z) = P(\zeta < z) = \iint_{x+y < z} f(x, y) \, dx \, dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-y} f(x, y) \, dx \right] dy$$

$$f_{\zeta}(z) = \int_{-\infty}^{\infty} f(z-y, y) \, dy = \int_{-\infty}^{\infty} f(x, z-x) \, dx$$

二 n 维随机变量

$$F(x_1, x_2, \dots, x_n) = P(\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_n < x_n)$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \dots dx_n$$

$$F_{\xi_i}(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{x_i} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) \, dx_1 \, dx_2 \dots dx_n$$

$$= P(\xi_1 < \infty, \xi_2 < \infty, \dots, \xi_i < x_i, \dots, \xi_n < \infty)$$

$$f(x_1, x_2, \dots, x_n) = \frac{\partial F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

§ 随机变量的数字特征

数字特征：用来描述随机变量概率分布主要特征的数量指标

数学期望 方差 协方差

一 数学期望

设离散型随机变量 ξ 可能取的值为 $x_1, x_2, \dots, x_i, \dots$ ，取这些值的概率分别为 $p_1, p_2, \dots, p_i, \dots$ ，若 $\sum_{i=1}^{\infty} x_i p_i < \infty$ ，则称

$$E(\xi) = \sum_{i=1}^{\infty} x_i p_i$$

为随机变量 ξ 的数学期望。

连续型随机变量的数学期望

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

运算法则

$$\textcircled{1} E(c) = c;$$

$$\textcircled{2} E\left(\sum_{i=1}^n a_i \xi_i\right) = \sum_{i=1}^n a_i E(\xi_i) \quad .$$

概率统计基础

函数 $\eta=\varphi(\xi)$ 的数学期望

$$E(\eta) = E[\varphi(\xi)] = \int_{-\infty}^{\infty} \varphi(x) f(x) dx$$

n 维随机变量的数学期望

$$E(\xi_i) = \int \cdots \int x_i f(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n$$

$$E(\xi) = (E(\xi_1), E(\xi_2), \cdots, E(\xi_n))$$

条件数学期望就是条件分布的数学期望

$$E(\xi_i | \xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_j = x_j, \cdots, \xi_n = x_n)$$

$$= \int_{-\infty}^{\infty} x_i f(x_i | x_1, x_2, \cdots, x_j, \cdots, x_n) dx_i, (i \neq j)$$

$n=2,$

$$E(\eta | \xi = x) = \int_{-\infty}^{\infty} y f(y|x) dy$$

$$E(\xi | \eta = y) = \int_{-\infty}^{\infty} x f(x|y) dx$$

ξ, η 相互独立,

$$E(\eta | \xi = x) = E(\eta)$$

$$E(\xi | \eta = y) = E(\xi)$$



二 方差

若随机变量对其数学期望偏差平方的数学期望存在，则称它为随机变量的**方差**，记为 $D(\xi)=E[\xi-E(\xi)]^2$

方差的平方根称为随机变量的**标准差**或**均方误差**，记为

$$\sigma(\xi) = \sqrt{D(\xi)}$$

离散型随机变量 $D(\xi) = \sum_{i=1}^{\infty} [x_i - E(\xi)]^2 p_i$

连续型随机变量 $D(\xi) = \int_{-\infty}^{\infty} [x - E(\xi)]^2 f(x) dx$

性质: ① $D(c) = 0$;

② $D(c\xi) = c^2 D(\xi)$;

③ $D\left(\sum_{i=1}^n c_i \xi_i\right) = \sum_{i=1}^n \sum_{j=1}^n c_i c_j E[(\xi_i - E(\xi_i))(\xi_j - E(\xi_j))]$ 。

常用公式 $D(\xi) = E(\xi^2) - [E(\xi)]^2$;

标准化变换 $T = \frac{\xi - E(\xi)}{\sqrt{D(\xi)}}$

n维随机变量的条件方差 (剩余方差)

$$\begin{aligned} D(\xi_i | \xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_j = x_j, \cdots, \xi_n = x_n) \\ = E[\xi_i - E(\xi_i | \xi_1 = x_1, \xi_2 = x_2, \cdots, \xi_j = x_j, \cdots, \xi_n = x_n)]^2 \\ = \int_{-\infty}^{\infty} [x_i - m_{\xi_i}(x_1, x_2, \cdots, x_j, \cdots, x_n)]^2 f(x_i | x_1, x_2, \cdots, x_j, \cdots, x_n) dx_i, (i \neq j) \end{aligned}$$

三 协方差和相关系数

协方差是表示两个随机变量之间相互关系密切程度的量, 定义为

$$\text{cov}(\xi, \eta) = E\{[\xi - E(\xi)][\eta - E(\eta)]\}$$

ξ, η 相互独立 $\text{cov}(\xi, \eta) = 0$

性质: ① $\text{cov}(\xi, \eta) = \text{cov}(\eta, \xi)$

② $\text{cov}(a\xi, b\eta) = ab \text{cov}(\xi, \eta)$, a, b 常数

③ $\text{cov}(\xi_1 + \xi_2, \eta) = \text{cov}(\xi_1, \eta) + \text{cov}(\xi_2, \eta)$

相关系数/标准协方差

$$\rho_{\xi\eta} = \frac{\text{cov}(\xi, \eta)}{\sqrt{D(\xi)D(\eta)}}$$

性质: ① $|\rho_{\xi\eta}| \leq 1$;

$\rho_{\xi\eta} > 0$, ξ 、 η 正相关 $\rho_{\xi\eta} < 0$, ξ 、 η 负相关;

ξ, η 的数学期望和方差为 $E(\xi)$ 、 $E(\eta)$ 、 $D(\xi)$ 、 $D(\eta)$, 则

$$\begin{aligned} & D\left[\frac{\xi - E(\xi)}{\sqrt{D(\xi)}} \pm \frac{\eta - E(\eta)}{\sqrt{D(\eta)}}\right] \\ &= E\left[\frac{\xi - E(\xi)}{\sqrt{D(\xi)}}\right]^2 + E\left[\frac{\eta - E(\eta)}{\sqrt{D(\eta)}}\right]^2 \pm 2E\left[\frac{\xi - E(\xi)}{\sqrt{D(\xi)}} \cdot \frac{\eta - E(\eta)}{\sqrt{D(\eta)}}\right] \\ &= 1 + 1 \pm 2\rho_{\xi\eta} \\ &\geq 0 \end{aligned}$$

有

$$-1 \leq \rho_{\xi\eta} \leq 1$$



② ξ 、 η 有线性关系, $|\rho_{\xi\eta}| = 1$;

$$\eta = a\xi + b$$

$$\begin{aligned}\text{cov}(\xi, \eta) &= E\{[\xi - E(\xi)][\eta - E(\eta)]\} \\ &= E\{[\xi - E(\xi)][a\xi + b - aE(\xi) - b]\} \\ &= aE[\xi - E(\xi)]^2 \\ &= aD(\xi)\end{aligned}$$

$$\begin{aligned}D(\eta) &= E[\eta - E(\eta)]^2 \\ &= E[a\xi + b - aE(\xi) - b]^2 \\ &= a^2 E[\xi - E(\xi)]^2 \\ &= a^2 D(\xi)\end{aligned}$$

$$\rho_{\xi\eta} = \frac{\text{cov}(\xi, \eta)}{\sqrt{D(\xi)D(\eta)}} = \frac{aD(\xi)}{|a|D(\xi)} = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

③ ξ, η 相互独立, $\rho_{\xi\eta} = 0$;

$\rho_{\xi\eta} = 0$, ξ, η 不一定相互独立。

矩： 统一的数字特征 —— 原点矩、中心矩、混合矩

k 阶原点矩 $E(\xi^k)$ $k=1$, 数学期望

k 阶中心矩 $E[\xi-E(\xi)]^k$ $k=2$, 方差

$k+l$ 阶混合矩 $E(\xi^k \eta^l)$

$k+l$ 阶中心混合矩 $E\{[\xi-E(\xi)]^k [\eta-E(\eta)]^l\}$ $k+l=2$, 协方差

n 维随机变量的方差： 每一分量的方差 + 任两分量的协方差

协方差矩阵

$$V = (v_{ij}) = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix}$$

其中 $v_{ij} = E\{[\xi_i - E(\xi_i)][\xi_j - E(\xi_j)]\}$

$$= \int \cdots \int [x_i - E(\xi_i)][x_j - E(\xi_j)] f(x_1, x_2, \cdots, x_n) dx_1 dx_2 \cdots dx_n$$

为 ξ_i, ξ_j 的协方差 $\text{cov}(\xi_i, \xi_j)$ 。

§ 几种常见概率分布

一 离散型随机变量分布

1. 二项分布

贝努里试验: $A \quad \bar{A}$ 相互独立

$$p \quad q = 1 - p$$

n 次试验中 A 出现的次数 k ξ ? $P(\xi=k)$

$$\begin{array}{l} n=4 \\ k=2 \end{array} \left\{ \begin{array}{lll} AA\bar{A}\bar{A} & A\bar{A}A\bar{A} & A\bar{A}\bar{A}A \\ \bar{A}A\bar{A}\bar{A} & \bar{A}\bar{A}A\bar{A} & \bar{A}\bar{A}\bar{A}A \end{array} \right. C_4^2 = 6$$

$$P_4(\xi=2) = C_4^2 p^2 q^{4-2}$$

$$P_n(\xi=k) = C_n^k p^k q^{n-k}$$

二项式 $(p+q)^n$ 的展开式中第 $k+1$ 项

性质: ① $P_n(\xi=k) \geq 0$

$$\text{②} \quad \sum_{k=0}^n P_n(\xi=k) = \sum_{k=0}^n C_n^k p^k q^{n-k} = 1$$

概率统计基础

分布函数 $F(x) = P(\xi < x) = \sum_{0 < k < x} C_n^k p^k q^{n-k}$

$$P(\xi = 1) = p$$

$$P(\xi = 0) = 1 - p$$

例: $n=10, p=0.2$, 求 $P(\xi \geq 2)$ 。

解: $P(\xi \geq 2) = 1 - P(\xi < 2) = 1 - \sum_{k=0}^1 C_n^k p^k q^{n-k}$

$$= 1 - (0.8^{10} + 10 \cdot 0.2 \cdot 0.8^9) = 0.625$$

数学期望 $E(\xi) = \sum_{k=0}^n k C_n^k p^k q^{n-k} = np$

方差 $D(\xi) = E(\xi^2) - E(\xi)^2 = np(1-p)$

2. 泊松分布

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \lambda > 0 \text{ 常数}$$

性质: ① $P_n(\xi = k) \geq 0$

② $\sum_{k=0}^{\infty} P_n(\xi = k) = 1$

概率统计基础

数学期望

$$E(\xi) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

方差

$$D(\xi) = \lambda$$

证:

$$E(\xi^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= e^{-\lambda} \lambda \sum_{k=0}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!}$$

$$= e^{-\lambda} \lambda \left[\sum_{k=0}^{\infty} \lambda \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right]$$

$$= \lambda^2 + \lambda$$

$$D(\xi) = E(\xi^2) - E(\xi)^2 = \lambda$$

$$\frac{P(k)}{P(k-1)} = \frac{\lambda}{k}$$

$$k < \lambda$$

$$P(k) \nearrow \quad k \nearrow$$

$$k > \lambda$$

$$P(k) \searrow \quad k \nearrow$$

二 连续型随机变量分布

1. 均匀分布

$$\xi \sim f(x)=c \quad a \leq x \leq b$$

分布密度函数性质 $\Rightarrow c = \frac{1}{b-a}$

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

$$P(\alpha < x < \beta) = \int_{\alpha}^{\beta} f(x) dx = \frac{\beta - \alpha}{b - a}$$

数学期望 $E(\xi) = \frac{1}{2}(b+a)$

方差 $D(\xi) = \frac{1}{12}(b-a)^2$

概率统计基础



$$E(\xi) = \int_{-\infty}^{\infty} xf(x) \mathrm{d} x$$

$$= \int_a^b x \frac{1}{b-a} \mathrm{d} x$$

$$= \frac{b^2 - a^2}{2(b-a)}$$

$$= \frac{1}{2}(b+a)$$

$$D(\xi) = E(\xi^2) - E(\xi)^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d} x - \left[\int_{-\infty}^{\infty} xf(x) \mathrm{d} x \right]^2$$

$$= \frac{b^3 - a^3}{3(b-a)} - \frac{1}{4}(b+a)^2$$

$$= \frac{1}{3}(b^2 + ba + a^2) - \frac{1}{4}(b^2 + 2ba + a^2)$$

$$= \frac{1}{12}(b-a)^2$$

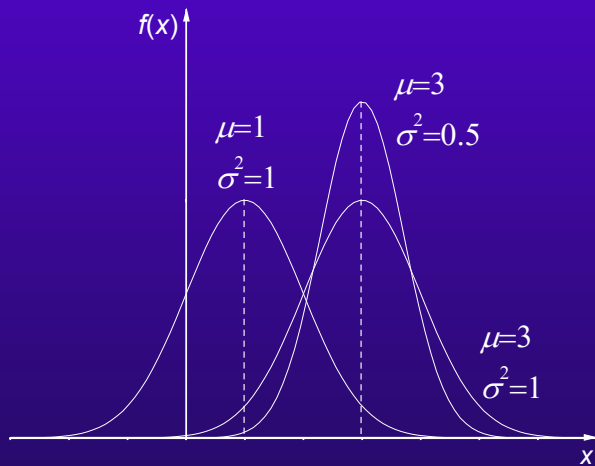
2. 指数分布

$$\xi \sim \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

3. 高斯分布（正态分布）

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



性质:

- ① 对称于 $x=\mu$, 且有极大值 $\frac{1}{\sigma\sqrt{2\pi}}$;
- ② 在 $x=\mu\pm\delta$ 处, 有拐点;
- ③ 固定 μ , $f(x)$ 形状随 σ 变;
固定 σ , $f(x)$ 位置随 μ 变。

概率统计基础

数学期望 $E(\xi) = \int_{-\infty}^{\infty} xf(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu$

方差 $D(\xi) = E[\xi - E(\xi)]^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$

$$\xi \sim N(\mu, \sigma^2) \leftrightarrow \text{正态分布}$$

$\mu=0, \sigma=1$ 标准正态分布

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

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} du$$

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

$$\Phi(u) = 1 - \Phi(-u) \quad u < 0$$

概率统计基础

例：已知 $\xi \sim N(1, 10^2)$ ，求 $P(\xi > 12)$ 及 $P(5 < \xi < 10)$ 。

解：① $P(\xi > 12) = 1 - P(\xi < 12) = 1 - F(12)$

$$= 1 - \Phi\left(\frac{12-1}{10}\right) = 1 - \Phi(1.1)$$

查正态分布表得 $\Phi(1.1) = 0.8643$

$$P(\xi > 12) = 1 - \Phi(1.1) = 1 - 0.8643 = 0.1352$$

$$\begin{aligned} \text{② } P(5 < \xi < 10) &= F(10) - F(5) = \Phi\left(\frac{10-1}{10}\right) - \Phi\left(\frac{5-1}{10}\right) \\ &= \Phi(0.9) - \Phi(0.4) = 0.8159 - 0.6554 = 0.1605 \end{aligned}$$

例：已知 $\xi \sim N(\mu, \sigma^2)$ ，求

$$\text{① } P(|\xi - \mu| < \sigma); \quad \text{② } P(|\xi - \mu| < 2\sigma); \quad \text{③ } P(|\xi - \mu| < 3\sigma)。$$

解：① $P(|\xi - \mu| < \sigma) = P(-\sigma < \xi - \mu < \sigma) = P\left(-1 < \frac{\xi - \mu}{\sigma} < 1\right)$

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6827;$$

$$\text{② } P(|\xi - \mu| < 2\sigma) = 2\Phi(2) - 1 = 0.9545;$$

$$\text{③ } P(|\xi - \mu| < 3\sigma) = 2\Phi(3) - 1 = 0.9973。$$

三 二维正态分布

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

数学期望 $E(\xi) = \mu_1 \quad E(\eta) = \mu_2$

方差 $D(\xi) = \sigma_1^2 \quad D(\eta) = \sigma_2^2$

边缘分布密度 $f_\xi(x) = \int_{-\infty}^{\infty} f(x, y) dy$

指数部分改为 $\left(\frac{y-\mu_2}{\sigma_2} - \rho \frac{x-\mu_1}{\sigma_1} \right)^2 + (1-\rho^2) \left(\frac{x-\mu_1}{\sigma_1} \right)^2$

$$f_\xi(x) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{y-\mu_2}{\sigma_2} - \rho \frac{x-\mu_1}{\sigma_1} \right)^2} dy$$

利用 $\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}$

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \quad f_\eta(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

概率统计基础

协方差

$$\text{cov}(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_1)(y - \mu_2) f(x, y) dx dy$$

$$= \rho \sigma_1 \sigma_2$$

$$\rho(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sigma_1 \sigma_2} = \rho$$

$$\rho = 0 \quad f(x, y) = \frac{1}{2\pi \sigma_1 \sigma_2} e^{-\frac{1}{2} \left[\frac{(x - \mu_1)^2}{\sigma_1^2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right]} = f_{\xi}(x) f_{\eta}(y)$$

例：已知 $\xi \sim N(0, 1)$, $\eta \sim N(0, 1)$, 求 $\zeta = \xi + \eta$ 的密度函数。

解： $F(z) = P(\zeta < z) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-y} f(x, y) dx \right] dy$

$$f(z) = \int_{-\infty}^{\infty} f(z - y, y) dy = \int_{-\infty}^{\infty} f_{\xi}(z - y) \cdot f_{\eta}(y) dy$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(z-y)^2/2} \cdot e^{-y^2/2} dy = \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{\infty} e^{-(y-\frac{z}{2})^2} dy = \frac{1}{2\sqrt{\pi}} e^{-z^2/4}$$

概率统计基础

用向量和矩阵表示二维正态分布

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \boldsymbol{\Sigma} = (\sigma_{ij}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^2 |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

$$\text{其中 } |\boldsymbol{\Sigma}| = \sigma_1^2 \sigma_2^2 (1 - \rho^2) \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{|\boldsymbol{\Sigma}|} \begin{bmatrix} \sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

n 维正态分布的概率密度函数

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

要点：① 熟悉正态分布密度的形式；

$$\varphi(x) = \frac{3}{\pi} e^{-(3x-5)^2/\pi}$$

② 正态变量的线性函数仍为正态变量；

$$\xi \sim N(\mu, \sigma^2) \quad \eta = k_1 \xi + k_2 \quad k_1 \neq 0$$

$$\eta \sim N(k_1 \mu + k_2, k_1^2 \sigma^2)$$

③ 独立正态变量的线性组合仍为正态变量；

$$\xi_i \sim N(\mu_i, \sigma_i^2)$$

$$\sum_{i=1}^n \alpha_i \xi_i \sim N\left(\sum_{i=1}^n \alpha_i \mu_i, \sum_{i=1}^n \alpha_i^2 \sigma_i^2\right)$$

$$\textcircled{4} \quad \xi, \eta \quad \varphi(x, y) = \begin{cases} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} + \frac{1}{2\pi} e^{-x^2} \cos x \cos y & |x| < \pi, |y| < \pi \\ \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} & \text{其它} \end{cases}$$

$\varphi(x, y)$ 不是二维正态分布。

例： 已知 $\xi_1 \sim N(1,2)$, $\xi_2 \sim N(0,3)$, $\xi_3 \sim N(2,1)$ 且 ξ_1, ξ_2, ξ_3 相互独立。求 $P(0 \leq 2\xi_1 + 3\xi_2 - \xi_3 \leq 6)$ 。

解： 令 $z = 2\xi_1 + 3\xi_2 - \xi_3$

$$E(z) = 0$$

$$\begin{aligned} D(z) &= 4D(\xi_1) + 9D(\xi_2) + D(\xi_3) \\ &= 8 + 27 + 1 = 36 = 6^2 \end{aligned}$$

$$P(0 \leq z \leq 6) = \Phi(1) - \Phi(0) = 0.3413$$

切比雪夫不等式：

$E(\xi) = \mu$, $D(\xi) = \sigma^2$, 则对 $\forall \varepsilon > 0$, 不等式

$$\begin{aligned} P(|\xi - \mu| \geq \varepsilon) &\leq \sigma^2 / \varepsilon^2 \\ P(|\xi - \mu| < \varepsilon) &\geq 1 - \sigma^2 / \varepsilon^2 \end{aligned} \quad \text{成立。}$$

证：

$$\begin{aligned} P(|\xi - \mu| \geq \varepsilon) &= \int_{|x - \mu| \geq \varepsilon} f(x) dx \\ &\leq \int \frac{|x - \mu|^2}{\varepsilon^2} f(x) dx \\ &= \sigma^2 / \varepsilon^2 \end{aligned}$$

例：事件 A 发生概率 $P(A)=0.5$ ，求 $n=1000$ 次试验中事件 A 发生的次数在400~600次之间的概率。

解： $\xi \sim B(1000, 0.5)$ $E(\xi) = 500$ $D(\xi) = 250$
 $P(400 \leq \xi \leq 600) = P(|\xi - 500| \leq 100)$
 $\geq 1 - 250/100^2 = 0.975$

§ 极限定理

一 随机序列的收敛

设 $\xi_1, \xi_2, \dots, \xi_n$ ， ξ 为随机变量或常数

$$\textcircled{1} \quad \forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} P(|\xi_n - \xi| > \varepsilon) = 0$$

不同于 $x_n \rightarrow a$ ，即 $|x_n - a| < \varepsilon$

$$\textcircled{2} \quad \xi_n \text{以概率1收敛} \quad \lim_{n \rightarrow \infty} P(\xi_n = \xi) = 1$$

$$\textcircled{3} \quad \xi_n \text{以均方收敛} \quad \lim_{n \rightarrow \infty} E(|\xi_n - \xi|^2) = 0$$

二 大数定理

1. 切比雪夫大数定理

设随机变量 $\xi_1, \xi_2, \dots, \xi_n$ 相互独立且同分布, 则对任给 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - \mu\right| < \varepsilon\right) = 1$$

若 $\xi_1, \xi_2, \dots, \xi_n$ 非同分布, 则有

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n E(\xi_i)\right| < \varepsilon\right) = 1$$

证: $\eta = \frac{1}{n} \sum_{i=1}^n \xi_i \quad E(\eta) = \mu \quad D(\eta) = \frac{1}{n} \sigma^2$

由切比雪夫不等式, 有 $P(|\eta - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$

$$1 \geq \lim_{n \rightarrow \infty} P(|\eta - \mu| < \varepsilon) \geq \lim_{n \rightarrow \infty} \left(1 - \frac{\sigma^2}{n\varepsilon^2}\right) = 1$$

2. 贝努利大数定理

设 n_A 为 n 次独立试验中 A 发生的次数, P 为 A 发生的概率, 对给定 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{n_A}{n} - P\right| < \varepsilon\right) = 1$$

证: 引入随机变量序列

$$\xi_k = \begin{cases} 1 & \text{第} k \text{次试验中} A \text{发生} \\ 0 & \text{第} k \text{次试验中} A \text{不发生} \end{cases}$$

则 $n_A = \sum_{k=1}^n \xi_k$, 且 ξ_k 相互独立, $E(\xi_k) = p$, $D(\xi_k) = p(1-p)$

由切比雪夫定理, 有 $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n \xi_k - p\right| < \varepsilon\right) = 1$

即 $\lim_{n \rightarrow \infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$



三 中心极限定理

1. 独立分布

设随机变量 $\xi_1, \xi_2, \dots, \xi_n$ 相互独立且同分布, $E(\xi) = \mu$,

$D(\xi) = \sigma^2$, 则对随机变量 $\eta = \frac{\sum \xi_i - n\mu}{\sqrt{n}\sigma}$ 有

$$\lim_{n \rightarrow \infty} P(\eta < y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt$$

若 $\xi_1, \xi_2, \dots, \xi_n$ 非同分布, $E(\xi_i) = \mu_i$, $D(\xi_i) = \sigma_i^2$, 则

对随机变量 $\eta = \frac{\sum \xi_i - \sum \mu_i}{\sqrt{\sum \sigma_i^2}}$ 有

$$\lim_{n \rightarrow \infty} P(\eta < y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt$$

2. 棣莫佛—拉普拉斯定理

若随机序列 $\{\xi_n\} (n=1,2,\dots)$ 为二项分布 $B(n,p)$ ，则对任意实数 x 有

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

n 的取值：如 $B(n,p)$ ，通常 n 不太大 ($n \leq 10$)， p 不太小 ($p \geq 0.1$)
 n 次独立试验中， $n_A \sim N(np, np(1-p))$

例： 保险公司有10000人参保，保险费18元 / 人。若投保人死亡，需赔偿2500元。问何种情况下保险公司亏本及其概率？（死亡率 $p=0.006$ 。）

解：
$$P(\xi > 72) = P\left(\frac{\xi - 60}{\sqrt{60(1-0.006)}} > 1.553862\right)$$

$$= 1 - P\left(\frac{\xi - 60}{\sqrt{60(1-0.006)}} \leq 1.553862\right) = 1 - 0.94 = 0.06$$

§ 样本及其分布

一 样本

总体：研究对象的全体 X

个体：组成总体的每一单元

样本：总体中若干个体组成的集合

样本应与总体同分布，其中的个体相互独立

为了反映总体性质，采用简单随机抽样的方法获得 $x_1, x_2, \dots, x_n$ 样本观测值，可以看作 n 维随机向量的观测值。这样的样本称为简单随机样本。

1. 经验分布函数

设 $x_1, x_2, \dots, x_n$ 为 X 的样本观测值， n 为样本容量，将它们按从大到小的顺序排列为 $x'_1, x'_2, \dots, x'_n$ ，定义

$$F_n(x) = \frac{x'_1, x'_2, \dots, x'_n \text{ 中小于 } x \text{ 的个数}}{n} = \begin{cases} 0 & x < x'_1 \\ \frac{k}{n} & x'_k < x < x'_{k+1} \\ 1 & x > x'_n \end{cases}$$

则称 $F_n(x)$ 为观测值 $x_1, x_2, \dots, x_n$ 的经验分布函数或样本分布函数。

2. 频率直方图（分布密度函数的近似求法）

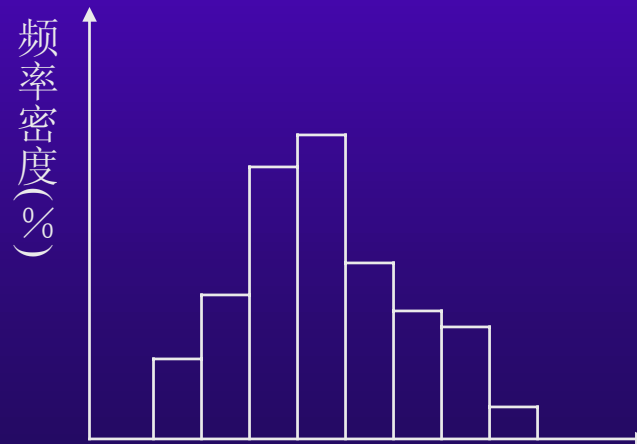
1) 整理样本观测值，并划分区间

排序 $x_1', x_2', \dots, x_n'$ ，并选取整数 $a \leq x_1', b \geq x_n'$ ；将 $[a, b]$ 分成 k 个互不相交的子区间 (x_i', x_{i+1}')

2) 计算各组的频率

落入每个子区间的个数 n_k (频数) 频率 $f = n_k/n$

3) 作频率直方图



频率直方图

二 统计量及其分布

1. 统计量的概念及常用的统计量

X	$X_1, X_2, \dots, X_n$	样本
	$g(X_1, X_2, \dots, X_n)$	样本函数 统计量

1) 样本均值
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

2) 样本方差
$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

3) 样本 k 阶原点矩
$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

4) 样本 k 阶中点矩
$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

2. 统计量的分布(抽样分布)

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n\mu = \mu$$

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

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= \frac{1}{n-1} E\left\{\sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2\right\} \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \mu)^2 - 2\sum_{i=1}^n (X_i - \mu)(\bar{X} - \mu) + \sum_{i=1}^n (\bar{X} - \mu)^2\right] \\ &= \frac{1}{n-1} \{n\sigma^2 - 2nE[(\bar{X} - \mu)^2] + nE[(\bar{X} - \mu)^2]\} \\ &= \frac{1}{n-1} \{n\sigma^2 - nE[(\bar{X} - \mu)^2]\} \\ &= \sigma^2 \end{aligned}$$

3. 若干统计量的分布

1) 正态总体样本均值的分布

设 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \dots, X_n$ 为它的一个样本, 则样本均值 $\bar{X} \sim N(\mu, \sigma^2/n)$

证: 若 $\xi \sim N(\mu, \sigma^2)$, 则 $\frac{\xi}{n} \sim N(\frac{\mu}{n}, \frac{\sigma^2}{n})$

而 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \sum_{i=1}^n \frac{X_i}{n}$ 为正态分布, 且

$$E(\bar{X}) = \mu \quad D(\bar{X}) = \frac{\sigma^2}{n}$$

即 $X \sim N(\mu, \sigma^2/n)$

2) χ^2 分布

设 $X \sim N(0, 1)$, $X_1, X_2, \dots, X_n$ 为它的随机样本, 则它们的平方和 $\chi^2 = \sum_{i=1}^n X_i^2$

服从参数为 n 的 χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$ 。

概率统计基础

概率密度函数

$$f(\chi^2, n) = \begin{cases} \frac{(\chi^2)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} & \chi^2 \geq 0 \\ 0 & \chi^2 < 0 \end{cases}$$

其中

$$\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx$$

- 性质:**
- (1) $E(\chi^2)=n$ $D(\chi^2)=2n$;
 - (2) χ^2 可加性 —— $\xi_1 \sim \chi^2(N_1), \xi_2 \sim \chi^2(N_2)$, 则
 $\xi_1 + \xi_2 \sim \chi^2(N_1 + N_2)$;
 - (3) $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \dots, X_n$ 为它的随机样本, 则
$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$
 - (4) $X_1, X_2, \dots, X_n$ 为正态分布 $N(\mu, \sigma^2)$ 的随机样本, 则
 - ① 样本均值 $\bar{X}$ 与样本方差 S^2 相互独立;
 - ② 统计量 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ 。

概率统计基础

证：①

$$\begin{aligned}\text{cov}(X_i - \bar{X}, \bar{X}) &= E[(X_i - \bar{X})(\bar{X} - \mu)] \\&= E\{[(X_i - \mu) - (\bar{X} - \mu)](\bar{X} - \mu)\} \\&= E[(X_i - \mu)(\bar{X} - \mu)] - E[(\bar{X} - \mu)^2] \\&= \frac{1}{n} E[(X_i - \mu)^2] + \frac{1}{n} \sum_{i \neq j} E[(X_i - \mu)(X_j - \mu)] \\&\quad - E[(\bar{X} - \mu)^2] \\&= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0\end{aligned}$$

$$\begin{aligned}\textcircled{2} \quad \frac{(n-1)S^2}{\sigma^2} &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 - n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2 \\ \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 &\sim \chi^2(n) \quad \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)\end{aligned}$$

约束条件 $\sum(X_i - \bar{X}) = 0$ 自由度减少一个

概率统计基础

统计量 $\frac{(n-1)S^2}{\sigma^2}$ 称为 χ^2 量

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = n-1$$

$$D\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$$

概率分布分位数

已知随机变量 X 的概率, 对于给定的 $\alpha (0 < \alpha < 1)$, 存在 x_α , 使得 $P(X > x_\alpha) = \alpha$, 则称 x_α 为上侧分位数。



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

$$S^2 = \frac{\chi^2 \sigma^2}{n-1}$$

$$E(S^2) = E\left(\frac{\chi^2 \sigma^2}{n-1}\right) = \frac{\sigma^2}{n-1} E(\chi^2) = \sigma^2$$

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

例：已知 $x \sim N(\mu, \sigma^2)$, $n=16$, 求 $P(S^2/\sigma^2 \leq 2.04)$ 。

解： $P(S^2 / \sigma^2 \leq 2.04) = P(15S^2 / \sigma^2 \leq 30.6)$
 $= P(\chi^2 \leq 30.6)$
 $= 1 - P(\chi^2 > 30.6)$
 $= 0.99$

3) t 分布

设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X, Y 相互独立, 则称随机变量

$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布, 并记为 $t \sim t(n)$ 。

概率密度函数

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(n/2)} (1 + t^2/n)^{-\frac{n+1}{2}} \quad -\infty < t < \infty$$

性质: (1) $E(t) = 0$ 对称性

$$D(t) = n/(n-2)$$

$$n \rightarrow \infty \quad t(n) \sim N(0,1) \quad \text{通常 } n > 45$$

(2) $X_1, X_2, \dots, X_n$ 为正态总体 $N(\mu, \sigma^2)$ 的随机样本, 则

$$\frac{(\bar{X} - \mu)}{S / \sqrt{n}} \sim t(n-1)$$

证:

$$\because \bar{X} \sim N(\mu, \sigma^2 / n)$$

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

$$\text{而 } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\text{且 } \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \text{ 与 } \frac{(n-1)S^2}{\sigma^2} \text{ 相互独立}$$

$$\text{有 } \frac{\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} / (n-1)}} = \frac{(\bar{X} - \mu)\sqrt{n}}{S} \sim t(n-1)$$

概率统计基础

(3) $X_1, X_2, \dots, X_n$ 和 $Y_1, Y_2, \dots, Y_n$ 分别为正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的随机样本, 且相互独立, 则

$$\textcircled{1} \quad \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

$$\textcircled{2} \quad \text{若 } \sigma_1^2 = \sigma_2^2 = \sigma^2, \text{ 则}$$
$$\frac{(\bar{X} - \mu_1) - (\bar{Y} - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2)$$

证: $\textcircled{1} \quad \because E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2 \quad D(\bar{X} - \bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$$\therefore \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

概率统计基础

$$\textcircled{2} \quad \therefore \frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1)$$

$$\frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1)$$

$$\therefore \frac{(n_1 - 1)S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

而
$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

故
$$\frac{(\bar{X} - \mu_1) - (\bar{Y} - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2)$$

4) F 分布

设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 并且 X 与 Y 相互独立, 则称随机变量 $F = \frac{\bar{X}/n_1}{\bar{Y}/n_2}$ 服从自由度为 (n_1, n_2) 的 F 分布, 记为

$$F = \frac{\bar{X}/n_1}{\bar{Y}/n_2} \sim F(n_1, n_2)$$

概率密度函数

$$f(z) = \begin{cases} \frac{\Gamma[(n_1 + n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \left(\frac{n_1}{n_2}\right) \left(\frac{n_1}{n_2} z\right)^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2} z\right)^{-\frac{n_1+n_2}{2}} & z \geq 0 \\ 0 & z < 0 \end{cases}$$

概率统计基础

性质: (1) $F_{1-\alpha}(n_1, n_2) = 1/F_{\alpha}(n_2, n_1)$
(2) $X_1, X_2, \dots, X_n$ 和 $Y_1, Y_2, \dots, Y_n$ 分别为正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的随机样本, 且相互独立, 则

$$\textcircled{1} \quad \frac{\sum (X_i - \mu_1)^2 / n_1 \sigma_1^2}{\sum (Y_i - \mu_2)^2 / n_2 \sigma_2^2} \sim F(n_1, n_2)$$

$$\because \begin{aligned} \sum (X_i - \mu_1)^2 / \sigma_1^2 &\sim \chi^2(n_1) \\ \sum (Y_i - \mu_2)^2 / \sigma_2^2 &\sim \chi^2(n_2) \end{aligned}$$

$$\textcircled{2} \quad \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

$$\because \begin{aligned} \frac{(n_1 - 1)S_1^2}{\sigma_1^2} &\sim \chi^2(n_1 - 1) \\ \frac{(n_2 - 1)S_2^2}{\sigma_2^2} &\sim \chi^2(n_2 - 1) \end{aligned}$$

概率统计基础

例：设 $X_1, X_2, \dots, X_7$ 为正态总体 $N(0, 0.5^2)$ 的样本，求：

$$(1) P\left(\sum_{i=1}^7 X_i^2 > 4\right); \quad (2) P\left(\sum_{i=1}^7 (X_i - \bar{X})^2 > 4\right)。$$

解：(1) $P\left(\sum_{i=1}^7 X_i^2 > 4\right) = P\left(\sum_{i=1}^7 \left(\frac{X_i - 0}{0.5}\right)^2 > \frac{4}{0.5^2}\right) = P(\chi^2(7) > 16) = 0.025$

$$(2) P\left(\sum_{i=1}^7 (X_i - \bar{X})^2 > 4\right) = P\left(\frac{(7-1)S^2}{0.5^2} > \frac{4}{0.5^2}\right) = P(\chi^2(6) > 16) = 0.014$$

例：从 $X \sim N(1000, \sigma^2)$ 中抽取容量为9的样本，得 $\bar{X}=940$ ， $S=100$ ，试求 $P(\bar{X} < 940)$ 。

解：

$$\begin{aligned} P(\bar{X} < 940) &= P\left(\frac{\bar{X} - 1000}{100 / \sqrt{9}} < \frac{940 - 1000}{100 / \sqrt{9}}\right) \\ &= P(t < -1.8) \\ &= P(t > 1.8) \\ &= 0.053 \end{aligned}$$

§ 参数估计

参数估计：用样本估计总体的某些特征值的方法。

一 参数的点估计

点估计：对总体的某些参数取值的估计。

通常总体 X 的分布函数的形式已知，但它含有一个或多个未知参数： $\theta_1, \theta_2, \dots, \theta_k$

1. 矩估计

矩估计法：利用样本矩来估计总体中相应的参数。

矩估计法的一般步骤：

$$\begin{aligned} (1) \quad a_r &= E(x^r) && \text{原点矩} \\ &= \int x^r f(x, \theta_1, \theta_2, \dots, \theta_k) dx \\ &= g(\theta_1, \theta_2, \dots, \theta_k) && r = 1, 2, \dots, k \end{aligned}$$

$$(2) \quad \theta_r = h(a_1, a_2, \dots, a_k)$$

$$\begin{aligned} (3) \quad A_r &= \frac{1}{n} \sum x_i^r && \text{样本原点矩} \\ \theta_r &= h(A_1, A_2, \dots, A_k) \end{aligned}$$

概率统计基础

例：求任意总体 X 的数学期望 μ 和方差 σ^2 的矩估计。

解： $\because a_1 = E(X) = \mu \quad a_2 = E(X^2) = D(X) + [E(X)]^2$

$$\therefore \mu = a_1 \quad \sigma^2 = a_2 - a_1^2$$

$$A_1 = \frac{1}{n} \sum X_i = \bar{X} \quad A_2 = \frac{1}{n} \sum X_i^2$$

$$\hat{\mu} = A_1 = \bar{X} \quad \hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n} \sum (X_i - \bar{X})^2 = S'^2$$

例： $X \sim B(m, p)$, $X_1, X_2, \dots, X_n$ 为它的样本, $x_1, x_2, \dots, x_n$ 为样本观测值, 求 m, p 的矩估计。

解： $\because a_1 = E(X) = mp \quad a_2 = E(X^2) = D(X) + [E(X)]^2$
 $= mp(1-p) + m^2 p^2$

$$\therefore m = \frac{a_1^2}{a_1 + a_1^2 - a_2} \quad p = \frac{a_1 + a_1^2 - a_2}{a_1}$$

$$\hat{m} = \frac{\bar{X}^2}{\bar{X} - S'^2} \quad \hat{p} = \frac{\bar{X} - S'^2}{\bar{X}}$$

2. 最大似然法

最大似然法：使观测结果具有最大概率的参数值作为未知参数的估计值的方法。

若总体 X 的分布密度函数为 $f(x, \theta_1, \theta_2, \dots, \theta_k)$ ，则定义样本 $X_1, X_2, \dots, X_n$ 的联合概率密度函数为似然函数

$$L(\theta) = L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(x_i, \theta_1, \theta_2, \dots, \theta_k)$$

要使似然函数 $L(\theta)$ 达到最大，必须满足

$$\frac{\partial}{\partial \theta_r} L(\theta) = 0 \quad r = 1, 2, \dots, k$$

由于 $L(\theta)$ 与 $\ln L(\theta)$ 的单调性相同，所以 $L(\theta)$ 与 $\ln L(\theta)$ 有相同的极值点。为了计算方便，常将似然函数 $L(\theta)$ 的极值点转化为 $\ln L(\theta)$ 的极值点。求解似然方程

$$\frac{\partial}{\partial \theta_r} \ln L(\theta) = 0 \quad r = 1, 2, \dots, k$$

从中得到参数 θ 的极大似然估计值。

求极大似然函数估计值的一般步骤:

- (1) 写出似然函数;
- (2) 对似然函数取对数, 并整理;
- (3) 求导数;
- (4) 解似然方程。

例: 设总体 X 的分布密度函数为 $f(X, \lambda) = e^{-\lambda} \lambda^X / X!$, $X_1, X_2, \dots, X_n$ 为 X 的样本, 求参数 λ 的最大似然估计值。

解:

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \lambda^{X_i} / X_i!$$

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^n X_i \ln \lambda - \sum_{i=1}^n \ln X_i$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = -n + \sum_{i=1}^n X_i / \lambda = 0$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

概率统计基础

例： 设总体 X 的分布密度函数为 $f(X, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$, $X_1, X_2, \dots, X_n$ 为 X 的样本, 求参数 μ, σ^2 的最大似然估计值。

解：
$$L(\mu, \sigma^2) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \right] = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2}}$$

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

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) = 0$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (X_i - \mu)^2 = 0$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

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

3. 估计量的衡量标准

例：设总体 X 的分布密度函数为 $f(X, \theta) = \begin{cases} (\theta+1)X^\theta & 0 < X < 1 \\ 0 & \text{其它} \end{cases}$ ， $X_1, X_2, \dots, X_n$ 为 X 的样本，求参数 θ 的估计值。

(1) 矩估计

$$a_1 = E(X) = \int_0^1 X(\theta+1)X^\theta dX = \frac{\theta+1}{\theta+2} = \bar{X}$$

$$\hat{\theta} = \frac{2\bar{X}-1}{1-\bar{X}}$$

(2) 最大似然估计

$$L(\theta) = \prod_{i=1}^n (\theta+1)X_i^\theta = (\theta+1)^n \left(\prod_{i=1}^n X_i\right)^\theta$$

$$\ln L(\theta) = n \ln(\theta+1) + \theta \sum_{i=1}^n \ln X_i$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta+1} + \sum_{i=1}^n \ln X_i = 0$$

$$\hat{\theta} = -1 - \frac{n}{\sum_{i=1}^n \ln X_i}$$

应该选用哪一种估计量？用什么标准来评价一个估计量的好坏？

常用标准：(1) 无偏性；(2) 有效性；(3) 一致性。

1) 无偏性

设 $X_1, X_2, \dots, X_n$ 是总体 X 的样本， $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 是总体参数 θ 的估计量。若 $E(\hat{\theta})$ 存在且 $E(\hat{\theta}) = \theta$ ，则称 $\hat{\theta}$ 为 θ 的无偏估计量。

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \mu$$

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

$$E(S'^2) = E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{n-1}{n} \sigma^2$$

$$\lim_{n \rightarrow \infty} E(S'^2) = \sigma^2$$

渐近无偏估计 $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$

2) 有效性

设 $\hat{\theta}_1(X_1, X_2, \dots, X_n)$ 与 $\hat{\theta}_2(X_1, X_2, \dots, X_n)$ 均为总体参数 θ 的无偏估计量, 若 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。

$$\bar{X} = \frac{1}{n} \sum X_i$$

$$Y = \sum c_i X_i \quad c_i > 0 \quad \sum c_i = 1$$

$$D(\bar{X}) < D(Y)$$

3) 一致性

若估计量 $\hat{\theta}$ 依概率收敛于总体参数 θ , 即对 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

则称 $\hat{\theta}$ 为 θ 的一致估计量。

样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 是 σ^2 的一致估计量

$$\because E(S^2) = \sigma^2 \quad D(S^2) = 2\sigma^4 / (n-1)$$

$$\therefore 1 \geq P(|S^2 - \sigma^2| < \varepsilon) \geq 1 - D(S^2) / \varepsilon^2 = 1 - 2\sigma^4 / [(n-1)\varepsilon^2] \rightarrow 1$$

二 参数的区间估计

区间估计：用随机区间来估计总体参数的可能取值范围

1. 置信水平和置信区间

设 $X_1, X_2, \dots, X_n$ 是总体 X 的样本， θ 为总体参数，对于给定的 $\alpha (0 < \alpha < 1)$ ，若存在两个统计量 $\hat{\theta}_1(X_1, X_2, \dots, X_n)$ 与 $\hat{\theta}_2(X_1, X_2, \dots, X_n)$ ，使得 $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$ ，则称随机区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 为参数 θ 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**置信水平**。

2. 区间估计的求法

常用步骤：(1) 寻找样本函数 $g(X_1, X_2, \dots, X_n, \theta)$ ；

(2) 给定 $1 - \alpha$ ，定出 a, b ，满足

$$P(a < g(X_1, X_2, \dots, X_n, \theta) < b) = 1 - \alpha;$$

(3) 解出 $\hat{\theta}_1, \hat{\theta}_2$ ，得到置信区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 。

1) 单体正态分布期望值 μ 的置信区间

$$X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2) \quad \hat{\mu} = \bar{X}$$

① σ^2 已知

$$u = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

$$P(-u_\alpha < u < u_\alpha) = 1 - \alpha$$

$$P\left(\left|\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}\right| < u_\alpha\right) = 1 - \alpha$$

$$P(\bar{X} - u_\alpha \sigma / \sqrt{n} < \mu < \bar{X} + u_\alpha \sigma / \sqrt{n}) = 1 - \alpha$$

② σ^2 未知

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$$

$$P(-t_\alpha < \frac{\bar{X} - \mu}{S / \sqrt{n}} < t_\alpha) = 1 - \alpha$$

$$\bar{X} - t_\alpha S / \sqrt{n} < \mu < \bar{X} + t_\alpha S / \sqrt{n}$$

2) 单体正态分布方差 σ^2 的置信区间

① μ 已知

$$\sum \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

$$P(\chi_1^2 < \frac{\sum (X_i - \mu)^2}{\sigma^2} < \chi_2^2) = 1 - \alpha$$

$$\therefore P(\chi^2 > \chi_{\alpha}^2) = \alpha$$

$$\therefore P(\chi^2 > \chi_{\alpha/2}^2) = \alpha/2 \quad P(\chi^2 > \chi_{1-\alpha/2}^2) = 1 - \alpha/2$$

$$\therefore P(\chi_{1-\alpha/2}^2 < \chi^2 < \chi_{\alpha/2}^2) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha$$

$$\frac{\sum (X_i - \mu)^2}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{\sum (X_i - \mu)^2}{\chi_{1-\alpha/2}^2}$$

② μ 未知

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{(n-1)S^2}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}$$

3) 非正态总体参数的区间估计

若总体 X 的分布未知，但样本容量很大，由中心极限定理，可近似地视 $\bar{X} \sim N(\mu, \sigma^2 / n)$ 。

例：已知 X 服从参数为 p 的二项分布，样本为 $X_1, X_2, \dots, X_n (n > 50)$ ，求参数 p 的置信度为 $1-\alpha$ 的置信区间。

解： $\sum_{i=1}^n X_i \sim N(np, np(1-p))$

$$\Rightarrow \frac{\bar{X} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

$$P\left(\left|\frac{\bar{X} - p}{\sqrt{p(1-p)/n}}\right| < u_{\alpha}\right) \approx 1 - \alpha$$

$$0 < \frac{(\bar{X} - p)^2}{p(1-p)/n} < u_{\alpha}^2$$

$$\text{即 } (n + u_{\alpha}^2)p^2 - (2n\bar{X} + u_{\alpha}^2)p + n\bar{X}^2 < 0$$

4) 两个正态总体的情形

(1) σ_1^2, σ_2^2 已知, $\mu_1 - \mu_2$ 的置信区间

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \sim N(0, 1)$$

$\mu_1 - \mu_2$ 的置信区间

$$\left((\bar{X} - \bar{Y}) - u_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{X} - \bar{Y}) + u_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

(2) σ_1^2, σ_2^2 未知但 $\sigma_1^2 = \sigma_2^2 = \sigma^2$, $\mu_1 - \mu_2$ 的置信区间

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2)$$

$\mu_1 - \mu_2$ 的置信区间

$$\left((\bar{X} - \bar{Y}) \pm t_\alpha \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right)$$

概率统计基础

(3) σ_1^2, σ_2^2 未知但 $n_1, n_2 > 50$, $\mu_1 - \mu_2$ 的置信区间

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_1^2 / n_1 + S_2^2 / n_2}} \sim N(0, 1)$$

$\mu_1 - \mu_2$ 的置信区间

$$\left((\bar{X} - \bar{Y}) - u_\alpha \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, (\bar{X} - \bar{Y}) + u_\alpha \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

(4) σ_1^2, σ_2^2 未知, $n_1 = n_2$, $\mu_1 - \mu_2$ 的置信区间

$$Z_i = X_i - Y_i$$

$$Z \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$\bar{Z} = \bar{X} - \bar{Y}$$

$$S_z^2 = \frac{1}{n} \sum_{i=1}^n [(X_i - Y_i) - (\bar{X} - \bar{Y})]^2$$

$\mu_1 - \mu_2$ 的置信区间

$$(\bar{Z} - t_\alpha S_z / \sqrt{n}, \bar{Z} + t_\alpha S_z / \sqrt{n})$$

(5) $\mu_1 - \mu_2$ 未知, 方差比 σ_1^2 / σ_2^2 的置信区间

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

σ_1^2 / σ_2^2 的置信区间

$$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{\alpha/2}}, \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\alpha/2}} \right)$$

(6) $\mu_1 - \mu_2$ 已知, 方差比 σ_1^2 / σ_2^2 的置信区间

$$\frac{\sum (X_i - \mu_1) / n_1 \sigma_1^2}{\sum (Y_j - \mu_2) / n_2 \sigma_2^2} \sim F(n_1, n_2)$$

σ_1^2 / σ_2^2 的置信区间

$$\left(\frac{1}{F_{\alpha/2}} \cdot \frac{n_2}{n_1} \cdot \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)}{\sum_{j=1}^{n_2} (Y_j - \mu_2)}, \frac{1}{F_{1-\alpha/2}} \cdot \frac{n_2}{n_1} \cdot \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)}{\sum_{j=1}^{n_2} (Y_j - \mu_2)} \right)$$



§ 假设检验

一 统计假设和检验原理

- 假设检验：指施加于一个或多个总体的概率分布或参数的假设。所作的假设可以是正确的，也可以是错误的。为判断所作的假设是否正确，从总体中抽取样本，根据样本的取值，按一定的原则进行检验，然后，作出接受或拒绝所作假设的决定。
- 假设检验的内容
 - 参数检验
 - 总体均值、均值差的检验
 - 总体方差、方差比的检验
 - 非参数检验
 - 分布拟合检验
 - 符号检验
 - 秩和检验

- 假设检验的理论依据

- * 反证法思想 假定成立
- * 小概率原理 小概率事件在一次试验中几乎不可能发生

- 假设检验的步骤

- ① 提出原假设 H_0 ;
- ② 选取统计量 λ ;
- ③ 根据显著水平 α , 决定拒绝域;
- ④ 计算 λ , 接受或拒绝 H_0 。

例: 某厂生产的螺钉,按标准强度为68克/mm², 而实际生产的螺钉强度 X 服从 $N(\mu, \sigma^2)$ 。若 $E(X) = \mu = 68$, 则认为这批螺钉符合要求, 否则认为不符合要求。为此提出如下假设: $H_0: \mu = 68$, 若显著水平为 α , 问何种情况下拒绝原假设 H_0 ?

解: $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \quad P\left(\left|\frac{\bar{X} - 68}{\sigma / \sqrt{n}}\right| > u_\alpha\right) = \alpha$

$$\bar{X} \in (-\infty, 68 - u_\alpha \sigma / \sqrt{n}] \cup [68 + u_\alpha \sigma / \sqrt{n}, +\infty) \quad \text{拒绝 } H_0$$

易犯错误: (1) $P(\text{拒绝}H_0|H_0\text{为真}) = \alpha$
(2) $P(\text{接受}H_0|H_0\text{为假}) = \leq$

二 参数的显著性检验

1. 总体期望的 u 检验 (方差已知)

1) 单正态总体均值的检验

原假设 $H_0 : E(X) = \mu$

大容量样本 $u = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

拒绝域 $|u| > u_\alpha$

2) 两正态总体均值相等的检验

原假设 $H_0 : \mu_1 = \mu_2$

样本 $\bar{X} \sim N(\mu_1, \sigma_1^2 / n_1) \quad \bar{Y} \sim N(\mu_2, \sigma_2^2 / n_2)$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \sim N(0,1)$$

2. 总体期望的 t 检验 (方差未知)

1) 单正态总体均值的检验

原假设 $H_0 : E(X) = \mu$

样本 $\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$

2) 两正态总体均值相等的检验

原假设 $H_0 : \mu_1 = \mu_2$

样本 $\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2)$

3. 单正态总体方差的 χ^2 检验

原假设 $H_0 : \sigma^2 = \sigma_0^2$

样本 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$

拒绝域 $\chi^2 > \chi_{\alpha/2}^2(n-1)$ 或 $\chi^2 < \chi_{1-\alpha/2}^2(n-1)$

4. 两正态总体方差相等的 F 检验

原假设 $H_0: \sigma_1^2 = \sigma_2^2$

样本 $F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$

拒绝域 $F > F_{\alpha/2}(n-1)$ 或 $F < F_{1-\alpha/2}(n-1)$

三 非参数检验

1. 分布函数的 χ^2 拟合检验

原假设 $H_0: F(X) = F_0(X)$ 假设分布

$$p_i = F_0(X_{i+1}) - F_0(X_i) = \int_{X_i}^{X_{i+1}} f(x) dx$$

$$S = \sum C_i (\omega_i - p_i)^2 \quad \omega_i: \text{出现频率}$$

卡尔·皮尔逊

$$C_i = n / p_i$$

$$S = \sum n(\omega_i - p_i)^2 / p_i = \sum (f_i - np_i)^2 / np_i$$

f_i : 观测频数 np_i : 理论频数

$$S \sim \chi^2(k - r - 1) \quad r: \text{未知参数个数} \quad k: \text{区间个数}$$

概率统计基础

定理：若 n 充分大(> 50), 则不论总体为何分别, 当 H_0 为真时, 统计量

$$\chi^2 = \sum (f_i - np_i)^2 / np_i \sim \chi^2(k-r-1)$$

$$P(\chi^2 > \chi_\alpha^2(k-r-1)) = \alpha$$

若 $\chi^2 > \chi_\alpha^2(k-r-1)$, 拒绝 H_0 ; 若 $\chi^2 < \chi_\alpha^2(k-r-1)$, 接受 H_0 。

例：每隔7s观测一次铀放射的 α 粒子数 x , 若观测100次

粒子数(i)	0	1	2	3	4	5	6	7	8	9	10	11
次数(f_i)	1	5	16	17	26	11	9	9	2	1	2	1

从理论上已知 $P(x=i) = \frac{e^{-\lambda} \lambda^i}{i!}$, 问假设的分布是否符合实际?

解： $H_0 : p(x=i) = p_0(x=i) = \frac{e^{-\lambda} \lambda^i}{i!}$

$$\hat{\lambda} = \bar{X} = \sum f_i i / \sum f_i = 4.2$$

$$p_i = \frac{e^{-4.2} 4.2^i}{i!} \quad \text{理论频数} \quad np_i = 100 \times \frac{e^{-4.2} 4.2^i}{i!}$$

概率统计基础



i	f_i		np_i		$f_i - np_i$	$(f_i - np_i)^2$	$(f_i - np_i)^2 / np_i$
0	1	6	1.50	7.80	-1.80	3.24	0.42
1	5		6.30				
2	16		13.23		2.77	7.67	0.58
3	17		18.52		-1.52	2.31	0.12
4	26		19.44		6.56	43.03	2.21
5	11		16.33		-5.33	28.41	1.74
6	9		11.43		-2.43	5.90	0.52
7	9		6.86		2.14	4.58	0.67
8	2	6	3.60	6.28	-0.28	0.08	0.01
9	1		1.68				
10	2		0.71				
11	1		0.27				
$\chi^2 = \sum (f_i - np_i)^2 / np_i$						6.27	

理论频数 $np_i < 5$ 合并区间

$$\alpha = 0.05 \quad \chi^2_{\alpha}(8-1-1) = 12.59$$

χ^2 检验的步骤:

- ① 提出原假设 $H_0: F(X) = F_0(X)$ 或 $f(X) = f_0(X)$;
- ② 将实轴 $(-\infty, \infty)$ 分成 k 个互不相交的区间 (a_i, a_{i+1}) ;
- ③ 根据假设的理论分布, 计算理论频数;

$$p_i = P(a_i < X < a_{i+1}) = F_0(X_{i+1}) - F_0(X_i) = \int_{X_i}^{X_{i+1}} f(x) dx$$
- ④ 统计落入区间 (a_i, a_{i+1}) 的实际频数 f_i ;
- ⑤ 计算皮尔逊统计量 $\chi^2 = \sum (f_i - np_i)^2 / np_i$;
- ⑥ 选取显著水平 α , 查表得到 $\chi^2_{\alpha}(k-r-1)$;
- ⑦ 作出判断: 拒绝 H_0 或接受 H_0 。

2. 柯尔莫哥洛夫—斯米尔诺夫检验

1) 柯尔莫哥洛夫检验 (K检验)

检验经验分布是否服从某种理论分布

$$\text{经验分布 } F_n(X) = \begin{cases} 0 & X < X_1 \\ i/n & X_i < X < X_{i+1} \\ 1 & X > X_n \end{cases}$$

概率统计基础



原假设 $H_0: F_n(X) = F(X)$ $F(X)$ 理论分布
统计量 $D_n = \max_{-\infty < x < \infty} |F_n(X) - F(X)|$

$$\lim_{n \rightarrow \infty} P(D_n < \frac{y}{\sqrt{n}}) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k e^{-2y^2 k^2} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

$$P(D_n = 0) \rightarrow 1 \quad n \rightarrow \infty$$

检验理论分布为 $F(X)$ 的拒绝域 $D_n > D_{n,\alpha}$
其中 $D_{n,\alpha}$ 由给定的 α 及 n 查表得到。

2) 斯米尔诺夫检验 (C检验)

检验两个总体是否有相同的分布形式

原假设 $H_0: F_1(X) = F_2(X)$

统计量 $D_{n_1, n_2} = \max_{-\infty < x < \infty} |F_{n_1}(X) - F_{n_2}(X)|$

检验两个分布相同的拒绝域 $KD_{n_1, n_2} \geq m(n_1, n_2, \alpha)$

其中 K 是 n_1, n_2 的最小公倍数, $m(n_1, n_2, \alpha)$ 为满足

$P(KD_{n_1, n_2} \geq r) \leq \alpha$ 的 r 中最小整数。

3. 符号检验法

1) 单一样本的中位数检验

$X_1, X_2, \dots, X_n$ 来自总体 X 的样本, $x_1, x_2, \dots, x_n$ 为样本的观察值, x_m 表示某个指定的中位数。把 x_i ($i=1, 2, \dots, n$)与 x_m 进行比较, 若 $x_i > x_m$, 记结果为+号, 若 $x_i < x_m$, 记为-号, $x_i = x_m$ 时, 可以忽略不考虑, 最后比较+号和-号的差异程度, 判断样本观察值的随机性。由于是以中位数为标准, 只用+或-符号反映各观察值与中位数的大小变化, 所以称这种非参数检验方法为单一样本的中位数符号检验。

中位数是变量一系列取值中位置居中的数, 如果样本观察值是随机变化的, 则比中位数 x_m 大的数和比中位数小的数的个数, 应该比较接近, 一旦不是这样, 可能说明有系统性的因素在起着支配作用。关于检验的标准, 对单一样本中位数检验而言, 可以利用二项分布计算出现+号或-号的概率, 然后把它与给定的显著性水平 α 进行对照, 就可以做出判断。

2) 成对样本的符号检验

从总体 $F(x)$ 和 $G(y)$ 中分别抽取容量为 n 的两个样本 $X_1, X_2, \dots, X_n$ 与 $Y_1, Y_2, \dots, Y_n$ ，抽取样本的方法是，分别从 $F(x)$ 和 $G(y)$ 中各抽出一个单位， $x_1, x_2, \dots, x_n$ 与 $y_1, y_2, \dots, y_n$ 表示两个样本的观察值，欲检验假设 $H_0: F(x) = G(y)$ 。

如果 H_0 成立，应该有

$$P\{x_i - y_i > 0, i=1, 2, \dots, n\} = P\{y_i - x_i > 0, i=1, 2, \dots, n\}。$$

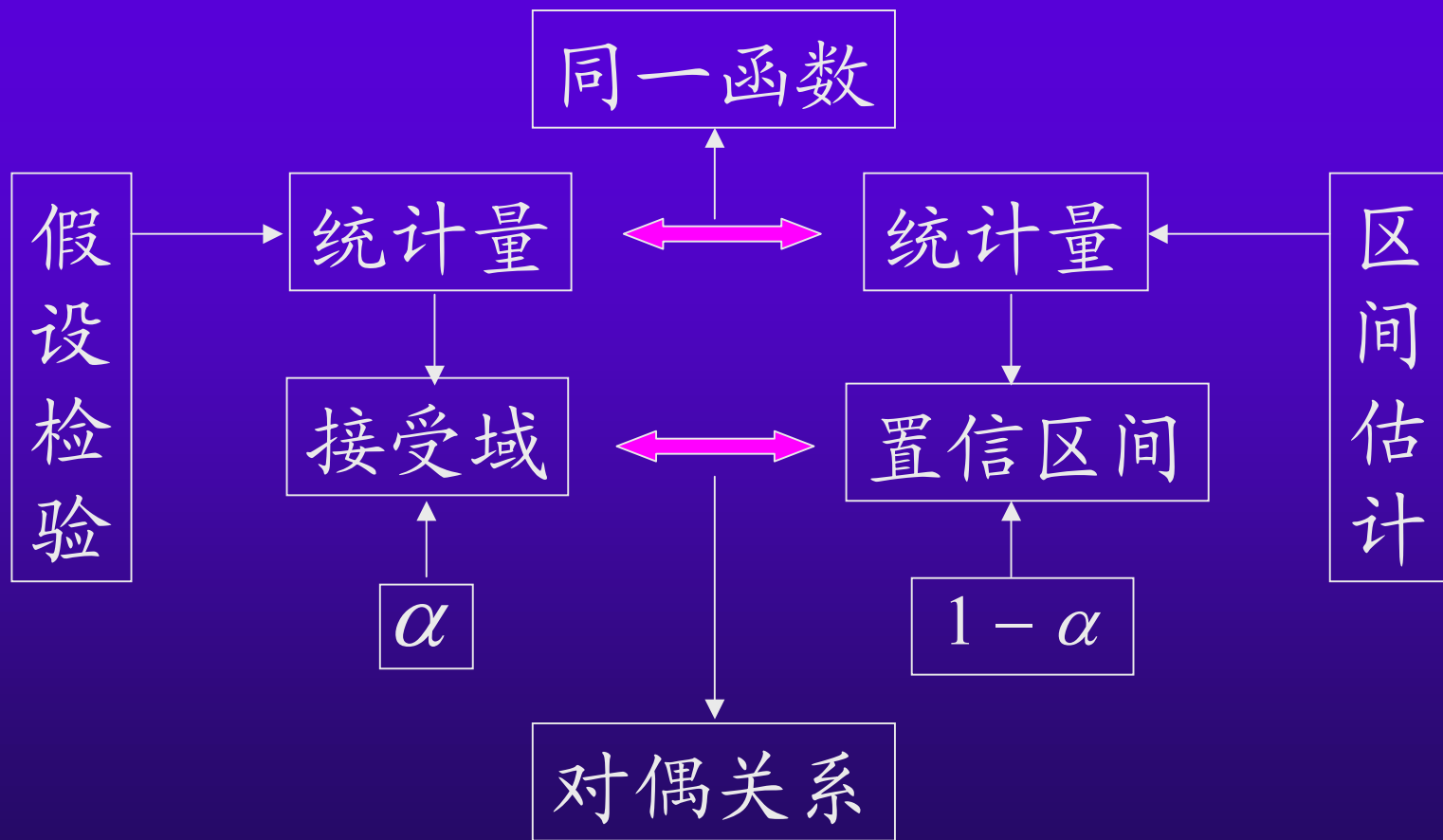
因此，可以用 $x_i - y_i > 0$ 或 $y_i - x_i > 0$ 发生的个数，作为假设的检验标准。

分别记 S_+ 和 S_- 为 $x_i - y_i > 0$ 和 $y_i - x_i > 0$ 的个数；对 $x_i - y_i = 0$ 的情况，可以不加考虑，或把 $x_i - y_i = 0$ 的个数一半取+号，一般取-号，如果是奇数个，减去1再作同样的处理，所以 $S_+ + S_- \leq n$ 。令 $S = \min(S_+, S_-)$ ，当 H_0 成立时，在显著性水平 α 下，有：

$$P(S \leq S_\alpha) = \alpha$$

式中， S_α 为成对符号检验临界值。此时，若 $S \leq S_\alpha$ ，则拒绝 H_0 ，反之，若 $S > S_\alpha$ ，则应该接受原假设 H_0 。

● 假设检验与区间估计的联系



§ 随机过程的基本知识

一 随机过程的基本定义

随机过程：兼有随机变量和函数的双重性质的过程。

$X(t)$ 是一个时间 t 的函数，而在时刻 t_1 又表现为随机变量
 $X(t)$ 这一随机过程可以是连续的，也可以是离散的
 称 $X(t)$ 的观测结果 $x_i(t)$ 为 $X(t)$ 的样本函数

二 随机过程的统计描述

1. 随机过程的概率分布

一维分布函数 $F_1(x, t) = P(X(t) < x) = \int_{-\infty}^x f_1(x, t) dx$

一维概率密度 $f_1(x, t)$

二维分布函数 $F_2(x_1, x_2, t_1, t_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_2(x_1, x_2, t_1, t_2) dx_1 dx_2$

分布函数族 $\{F_1(x, t), F_2(x_1, x_2, t_1, t_2), \dots, F_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)\}$

分布密度族 $\{f_1(x, t), f_2(x_1, x_2, t_1, t_2), \dots, f_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)\}$

2. 随机过程的数字特征

数学期望 $E(X(t)) = \mu_X(t) = (\mu_X(t_1), \mu_X(t_2), \dots, \mu_X(t_n))$

$$\mu_X(t_i) = E(X(t_i)) = \int_{-\infty}^{\infty} xf(x, t_i) dx \quad \text{集合平均}$$

方差 $D(X(t)) = D_X(t) = (D_X(t_1), D_X(t_2), \dots, D_X(t_n))$

$$D_X(t_i) = E\{[X(t_i) - \mu_X(t_i)]^2\}$$

协方差 $\text{COV}_X(t_i, t_j) = E\{[X(t_i) - \mu_X(t_i)][X(t_j) - \mu_X(t_j)]\}$

相关函数 $R_X(t_i, t_j) = E[X(t_i)X(t_j)]$

$$\begin{aligned} \text{COV}_X(t_i, t_j) &= E\{[X(t_i) - \mu_X(t_i)][X(t_j) - \mu_X(t_j)]\} \\ &= E[X(t_i)X(t_j)] - E[X(t_i)\mu_X(t_j)] \\ &\quad - E[X(t_j)\mu_X(t_i)] + E[\mu_X(t_i)\mu_X(t_j)] \\ &= E[X(t_i)X(t_j)] - E[X(t_i)]\mu_X(t_j) \\ &\quad - E[X(t_j)]\mu_X(t_i) + \mu_X(t_i)\mu_X(t_j) \\ &= R_X(t_i, t_j) - \mu_X(t_i)\mu_X(t_j) \end{aligned}$$

三 平稳随机过程

1. 定义

对任意的 τ 和 n ，都有下列等式成立：

$$\begin{aligned} f_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) \\ = f_n(x_1, x_2, \dots, x_n; t_1 + \tau, t_2 + \tau, \dots, t_n + \tau) \end{aligned}$$

则称随机过程 $X(t)$ 为**平稳随机过程**(严平稳过程)。

若 $X(t)$ 满足 $E(X(t)) = \mu_X(t) = \mu_X$

$$\text{COV}_{XX}(t_i, t_j) = \text{COV}(t_j - t_i)$$

则称 $X(t)$ 为**广义平稳过程**或**宽平稳过程**。

广义平稳过程 $\overset{\text{正态}}{\iff}$ 狭义平稳过程

2. 平稳过程的各态历经性

计算集合平均(如 μ_X R_{XX})有实际困难，期望以时间平均来代替集合平均

概率统计基础

时间平均 $\langle X(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt$

期望的时间平均有下面的性质：

① $\langle X(t) \rangle = E(X(t)) = \mu_X$ 期望各态历经

② $\langle X(t)X(t+\tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t+\tau) dt$
 $= E[X(t)X(t+\tau)] = R_X(\tau)$

自相关函数各态历经

若条件①②均满足，则称 $X(t)$ 是各态历经过程。

例：讨论随机过程 $X(t) = a \cos(\omega t + \theta)$ 的各态历经性， $\theta \in (0, 2\pi)$ 。

解： θ 的概率密度为 $f(\theta) = \begin{cases} 1/2\pi & 0 < \theta < 2\pi \\ 0 & \text{其它} \end{cases}$

概率统计基础

均值 $\mu_X(t) = \int_0^{2\pi} a \cos(\omega t + \theta) \frac{1}{2\pi} d\theta = 0$ 集合平均

$\langle X(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T a \cos(\omega t + \theta) dt = 0$ 时间平均

自相关函数 $R_X(t) = E[a \cos(\omega t_1 + \theta) \cdot a \cos(\omega t_2 + \theta)]$ 集合平均

$$= a^2 \int_0^{2\pi} \cos(\omega t_1 + \theta) \cos(\omega t_2 + \theta) \frac{1}{2\pi} d\theta$$

$$= \frac{a^2}{2} \cos \omega(t_1 - t_2) = \frac{a^2}{2} \cos \omega\tau$$

$\langle X(t_1)X(t_2) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t_1)x(t_2) dt$ 时间平均

$$= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T a^2 \cos(\omega t_1 + \theta) \cos(\omega t_2 + \theta) dt$$

$$= \frac{a^2}{2} \cos \omega\tau$$



3. 平稳随机过程相关函数的性质

1) 自相关函数

① $R_X(\tau) = R_X(-\tau)$ 偶函数

② $R_X(0) = \text{COV}_X(0) + \mu_X^2 > 0$

③ $|R_X(\tau)| \leq R_X(0) \quad |\text{COV}_X(\tau)| \leq \text{COV}_X(0)$

④ $R_X(\tau)$ 是非负定的, 即对任意数组 $t_1, t_2, \dots, t_n$ 和任意函

数 $g(t)$ 都有
$$\sum_{i,j=1}^n R_X(t_i - t_j) g(t_i) g(t_j) \geq 0$$

⑤ 若 $X(t)$ 和 $X(t+\tau)$ 在 $\tau \rightarrow \infty$ 时独立, 且 $E[X(t)] = 0$, 则

$$\lim_{\tau \rightarrow \infty} R_X(\tau) = 0$$

概率统计基础

$$\begin{aligned} E\{[X(t) + \lambda X(t + \tau)]^2\} \\ = E[X^2(t)] + 2\lambda E[X(t)X(t + \tau)] + \lambda^2 E[X^2(t + \tau)] \\ = R_X(0) + 2\lambda R_X(\tau) + \lambda^2 R_X(0) \geq 0 \end{aligned}$$

$$\begin{aligned} \sum_{i,j=1}^n R_X(t_i - t_j) g(t_i) g(t_j) &= \sum_{i,j=1}^n E[X(t_i)X(t_j)] g(t_i) g(t_j) \\ &= E\left\{ \sum_{i,j=1}^n [X(t_i)X(t_j) g(t_i) g(t_j)] \right\} \\ &= E\left\{ \left[\sum_{i,j=1}^n X(t_i) g(t_i) \right]^2 \right\} \geq 0 \end{aligned}$$

$$R_X(\tau) = \text{COV}_X(\tau) + E[X(t)]E[X(t + \tau)]$$

$$\tau \rightarrow \infty \quad \text{COV}_X(\tau) = 0$$

$$\lim_{\tau \rightarrow \infty} R_X(\tau) = \lim_{\tau \rightarrow \infty} E[X(t)]E[X(t + \tau)] = 0$$

2) 互相关函数

① $R_{XY}(-\tau) = R_{YX}(\tau)$ 非奇非偶

② $|R_{XY}(\tau)|^2 \leq R_X(0)R_Y(0)$

$$|\text{COV}_{XY}(\tau)|^2 \leq \text{COV}_X(0)\text{COV}_Y(0)$$

③ $R_{XY}(\tau) = \text{COV}_{XY}(\tau) + \mu_X\mu_Y$

平稳过程 $X(t)$ 满足 $X(t) = X(t+T)$, T 为周期, 则称 $X(t)$ 为周期过程, 且有 $R_X(t) = R_X(t+T)$ 。

耀斑 $\xrightarrow{20h}$ 磁暴, 试考虑它们的相关性。

发射信号 $X(t)$, 接受返回信号 $Y(t) = X(t-\tau) + \eta(t)$, $\eta(t)$ 为噪音

$$\begin{aligned} R_{XY}(\tau) &= E[X(t)Y(t)] \\ &= E[X(t)X(t-\tau)] + E[X(t)\eta(t)] \\ &= R_X(\tau) \end{aligned}$$