

概率论与数理统计总结

刘阳

2019 年 6 月 12 日

目录

1 第一章 概率论的基本概念	3
1.1 随机试验	3
1.2 样本空间、随机事件	3
1.2.1 样本空间	3
1.2.2 随机事件	3
1.2.3 事件间的关系与事件的运算	3
1.3 频率与概率	4
1.3.1 频率	4
1.3.2 概率	5
1.4 等可能概型（古典概型）	5
1.5 条件概率	6
1.5.1 条件概率	6
1.5.2 乘法定理	6
1.5.3 全概率公式和贝叶斯公式	6
1.6 独立性	7
2 第二章 随机变量及其分布	7
2.1 随机变量	7
2.2 离散型随机变量及其分布率	7
2.2.1 (0-1) 分布	7
2.2.2 伯努利试验、二项分布	8
2.2.3 泊松分布	8
2.3 随机变量的分布函数	8
2.4 连续型随机变量及其概率密度	8

目录	2
2.4.1 均匀分布	9
2.4.2 指数分布	9
2.4.3 正态分布	9
2.5 随机变量的函数的分布	10
3 第三章 多维随机变量及其分布	10
3.1 二维随机变量	10
3.2 边缘分布	12
3.3 条件分布	13
3.4 相互独立的随机变量	13
4 第四章 随机变量的数字特征	14
4.1 数学期望	14
4.2 方差	15
4.3 协方差及相关系数	15
4.4 矩、协方差矩阵	16
5 第五章 大数定律及中心极限定理	17
5.1 大数定律	17
5.2 中心极限定理	18
6 第六章 样本及抽样分布	18
6.1 随机样本	18
6.2 抽样分布	19
6.2.1 χ^2 分布	20
6.2.2 t 分布	20
6.2.3 F 分布	21
6.2.4 正态总体的样本均值与样本方差的分布	21
7 第七章 参数估计	22
7.1 点估计	22
7.1.1 矩估计法	22
7.1.2 最大似然估计法	23
7.2 估计量的评选标准	25
7.2.1 无偏性	25
7.2.2 有效性	25
7.2.3 相合性	25

1 第一章 概率论的基本概念	3
7.3 区间估计	26
7.4 正态总体均值与方差的区间估计	26
7.4.1 单个总体 $N(\mu, \sigma^2)$ 的均值 μ 的置信区间	26
7.4.2 两个总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ 的均值差 $\mu_1 - \mu_2$ 的置信区间	27

1 第一章 概率论的基本概念

1.1 随机试验

随机试验: 1. 可以在相同条件下重复地进行;
2. 每次试验的可能结果不止一个, 并且能事先明确实验的所有可能结果;
3. 进行一次实验之前不能确定哪一个结果会出现.

1.2 样本空间、随机事件

1.2.1 样本空间

随机试验 E 的所有可能结果组成的集合称为 **随机试验**.
样本空间的元素, 即 E 的每个结果称为 **样本点**.

1.2.2 随机事件

试验 E 的样本空间 S 的子集称为 **随机事件**, 简称 **事件**.
每次试验中, 当且仅当这一子集的一个样本点出现称为 **事件发生**.
有一个样本点组成的单点集称为 **基本事件**.
样本空间 S 包含所有的样本点, 它是 S 自身的子集, 在每次试验中它总是发生的, S 成为 **必然事件**.
空集 \emptyset 不包含任何样本点, 它也作为样本空间的子集, 他在每次试验中都不发生, \emptyset 称为 **不可能事件**.

1.2.3 事件间的关系与事件的运算

1. 若 $A \subset B$, 则称事件 B 包含事件 A , 这指的是事件 A 发生必导致事件 B 发生.
若 $A \subset B$ 且 $B \subset A$, 即 $A = B$, 则称事件 A 与事件 B 相等.

2. 事件 $A \cup B = \{x | x \in A \text{ 或 } x \in B\}$ 称为事件 A 与事件 B 的 **和事件**. 当且仅当 A, B 中至少一个发生时, 事件 $A \cup B$ 发生.

类似地, 称 $\bigcup_{k=1}^n A_k$ 为 n 个事件 A_1, A_2, \dots, A_n 的 **和事件**; 称 $\bigcup_{k=1}^{\infty} A_k$ 为可列个事件 A_1, A_2, \dots 的 **和事件**.

3. 事件 $A \cap B = \{x | x \in A \text{ 且 } x \in B\}$ 称为事件 A 与事件 B 的 **积事件**. 当且仅当 A, B 同时发生时, 事件 $A \cap B$ 发生. $A \cap B$ 也记作 AB .

类似地, 称 $\bigcap_{k=1}^n A_k$ 为 n 个事件 A_1, A_2, \dots, A_n 的 **积事件**; 称 $\bigcap_{k=1}^{\infty} A_k$ 为可列个事件 A_1, A_2, \dots 的 **积事件**.

4. 事件 $A - B = \{x | x \in A \text{ 且 } x \notin B\}$ 称为事件 A 与事件 B 的 **差事件**. 当且仅当 A 发生、 B 不发生时事件 $A - B$ 发生.

5. 若 $A \cap B = \emptyset$ 则称事件 A 与事件 B 是 **互不相容的**, 或 **互斥的**. 这指的是事件 A 与事件 B 不能同时发生. 基本事件也是两两互不相容的.

6. 若 $A \cup B = S$ 且 $A \cap B = \emptyset$, 则称事件 A 与事件 B 互为 **逆事件**. 又称事件 A 与事件 B 互为 **对立事件**. 这指的是对每次试验而言, 事件 A, B 中必有一个发生, 且仅有一个发生. A 的对立事件记为 \bar{A} . $\bar{A} = S - A$

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

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

德摩根律: $A \cup B = \overline{\bar{A} \cap \bar{B}}$; $A \cap B = \overline{\bar{A} \cup \bar{B}}$.

1.3 频率与概率

1.3.1 频率

在相同条件下, 进行了 n 次试验, 在这 n 次试验中, 事件 A 发生的次数为 n_A 称为事件 A 发生的 **频数**. 比值 $\frac{n_A}{n}$ 称为事件 A 发生的 **频率**, 并记成 $f_n(A)$

性质: 1. $0 \leq f_n(A) \leq 1$;

$$2. f_n(S) = 1;$$

3. 若 A_1, A_2, \dots, A_k 是两两互不相容的事件, 则

$$f_n(A_1 \cup A_2 \cup \dots \cup A_k) = f_n(A_1) + f_n(A_2) + \dots + f_n(A_k).$$

1.3.2 概率

设 E 是随机试验, S 是它的样本空间. 对于 E 的每一事件 A 赋予一个实数, 记为 $P(A)$, 称为事件 A 的 **概率**, 如果集合函数 $P(\cdot)$ 满足以下条件:

1. **非负性**: 对于每一个事件 A , 有 $P(A) \geq 0$;
2. **规范性**: 对于必然事件 S , 有 $P(S) = 1$;
3. **可列可加性**: 设 A_1, A_2, \dots 是两两互不相容的事件, 即对于 $A_i A_j = \emptyset$, $i \neq j, i, j = 1, 2, \dots$, 有

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots.$$

性质: 1. $P(\emptyset) = 0$.

2. 若 A_1, A_2, \dots, A_n 是两两互不相容的事件, 则有

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

3. 设 A, B 是两个事件, 若 $A \subset B$, 则有

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

$$P(B) \geq P(A).$$

4. 对于任一事件 A , 有

$$P(A) \leq 1.$$

5. 对于任意两事件 A, B 有

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

1.4 等可能概型 (古典概型)

1. 试验的样本空间只包含有限个元素.
2. 试验中每个基本事件发生的可能性相同.

具有以上两个特点的试验称为 **等可能概型**. 它在概率论发展初期曾是主要的研究对象, 所以也成为 **古典概型**.

设试验的样本空间为 $S = \{e_1, e_2, \dots, e_n\}$ 若 A 包含 k 个基本事件则有:

$$P(A) = \sum_{j=1}^k P(\{e_{i_j}\}) = \frac{k}{n} = \frac{A \text{ 包含的基本事件数}}{S \text{ 中基本事件的总数}}.$$

超几何分布的概率公式:

$$p = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}.$$

1.5 条件概率

1.5.1 条件概率

设 A, B 是两个事件, 且 $P(A) > 0$, 称

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

为在事件 A 发生下事件 B 发生的 **条件概率**.

1.5.2 乘法定理

设 $P(A) > 0$ 则有

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

1.5.3 全概率公式和贝叶斯公式

全概率公式: 设试验 E 的样本空间为 S , A 为 E 的事件, B_1, B_2, \dots, B_n 为 S 的一个划分, 且 $P(B_i) > 0 (i = 1, 2, \dots, n)$, 则

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n).$$

贝叶斯 (Bayes) 公式: 设试验 E 的样本空间为 S , A 为 E 的事件, B_1, B_2, \dots, B_n 为 S 的一个划分, 且 $P(A) > 0, P(B_i) > 0 (i = 1, 2, \dots, n)$, 则

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}, \quad i = 1, 2, \dots, n.$$

1.6 独立性

设 A, B 两事件如果满足等式

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

则称事件 A, B **相互独立**, 简称 A, B **独立**.

定理: 1. 设 A, B 是两事件, 且 $P(A) > 0$. 若 A, B 相互独立, 则 $P(B|A) = P(B)$. 反之亦然.

2. 若事件 A 与 B 相互独立, 则下列各对事件也相互独立: A 与 \bar{B} , \bar{A} 与 B , \bar{A} 与 \bar{B} .

一般, 设 A_1, A_2, \dots, A_n 是 $n(n \geq 2)$ 个事件, 如果对于其中任意 2 个, 任意 3 个, \dots , 任意 n 个事件的积事件的概率, 都等于各事件概率之积, 则称事件 A_1, A_2, \dots, A_n 相互独立.

推论: 1. 若事件 $A_1, A_2, \dots, A_n(n \geq 2)$ 相互独立, 则其中任意 $k(2 \leq k \leq n)$ 个事件也是相互独立的.

2. 若 n 个事件 $A_1, A_2, \dots, A_n(n \geq 2)$ 相互独立, 则将 A_1, A_2, \dots, A_n 中任意多个事件换成它们各自的对立事件, 所得的 n 个事件仍相互独立.

2 第二章 随机变量及其分布

2.1 随机变量

设随机试验的样本空间为 $S = \{e\}$. $X = X(e)$ 是定义在样本空间上的单值函数. 称 $X = X(e)$ 为 **随机变量**.

2.2 离散型随机变量及其分布率

2.2.1 (0-1) 分布

设随机变量 X 只可能取 0 与 1 两个值, 它的分布律是

$$P\{X = k\} = p^k(1-p)^{1-k}, k = 0, 1 \quad (0 < p < 1)$$

则称 X 服从以 p 为参数的 (0-1) 分布或 **两点分布**

2.2.2 伯努利试验、二项分布

设试验 E 只有两个可能结果: A 及 \bar{A} , 则称 E 为 **伯努利 (Bernoulli) 试验**.

将 E 独立重复地进行 n 次, 则称这一串重复的独立试验为 **n 重伯努利试验**.

以 X 表示 n 重伯努利试验中事件 A 发生的次数, 每次伯努利试验中 A 事件发生的概率为 p 称随机变量 X 服从参数为 n, p 的 **二项分布**, 并记为 $X \sim b(n, p)$. 它的分布律是

$$P\{x = k\} = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n$$

2.2.3 泊松分布

设随机变量 X 的所有可能取的值为 $0, 1, 2, \dots$, 而取各个值的概率为

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

其中 $\lambda > 0$ 是常数. 则称 X 服从参数为 λ 的 **泊松分布**, 记为 $X \sim \pi(\lambda)$

泊松定理: 设 $\lambda > 0$ 是一个常数, n 是任意正整数, 设 $np_n = \lambda$, 则对于任意固定的非负整数 k 有

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

一般, 当 $n \geq 20, p \leq 0.05$ 时用 $\frac{\lambda^k e^{-\lambda}}{k!} (\lambda = np)$ 作为 $\binom{n}{k} p^k (1-p)^{n-k}$ 的近似值效果颇佳.

2.3 随机变量的分布函数

设 X 是一个随机变量, x 是任意实数, 函数

$$F(x) = P\{X \leq x\}, -\infty < x < \infty$$

称为 X 的 **分布函数**.

2.4 连续型随机变量及其概率密度

如果对于随机变量 X 的分布函数 $F(x)$, 存在非负可积函数 $f(x)$, 使对于任意实数 x 有

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

则称 X 为 **连续型随机变量**, 其中函数 $f(x)$ 称为 X 的 **概率密度函数**, 简称 **概率密度**.

2.4.1 均匀分布

若连续型随机变量 X 具有概率密度

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

则称 X 在区间 (a, b) 上服从 **均匀分布**. 记为 $X \sim U(a, b)$

2.4.2 指数分布

若连续型随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0, \\ 0, & \text{其他}, \end{cases}$$

其中 $\theta > 0$ 为常数, 则称 X 服从参数为 θ 的 **指数分布**.

2.4.3 正态分布

若连续型随机变量 X 的概率密度为

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

其中 $\mu, \sigma (\sigma > 0)$ 为常数, 则称 X 服从参数为 μ, σ 的 **正态分布** 或 **高斯 (Gauss)** 分布, 记为 $X \sim N(\mu, \sigma^2)$.

性质: 1. 曲线关于 $x = \mu$ 对称. 这表明对于任意 $h > 0$ 有

$$P\{\mu - h < X \leq \mu\} = P\{\mu < X \leq \mu + h\}.$$

2. 当 $x = \mu$ 时取到最大值

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}.$$

特别, 当 $\mu = 0, \sigma = 1$ 时称随机变量 X 服从 **标准正态分布**. 其概率密度和分布函数分别用 $\varphi(x), \Phi(x)$ 表示, 即有

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

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

易知

$$\Phi(-x) = 1 - \Phi(x).$$

引理: 若 $X \sim N(\mu, \sigma^2)$, 则 $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

2.5 随机变量的函数的分布

设随机变量 X 具有概率密度 $f_X(x)$, $-\infty < x < \infty$, 又设函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$ (或恒有 $g'(x) < 0$). 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

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

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

3 第三章 多维随机变量及其分布

3.1 二维随机变量

一般, 设 E 是一个随机试验, 它的样本空间是 $S = \{e\}$, 设 $X = X(e)$ 和 $Y = Y(e)$ 是定义在 S 上的随机变量, 由它们构成的一个向量 (X, Y) , 叫做 **二维随机向量**或 **二维随机变量**.

设 (X, Y) 是二维随机变量, 对于任意实数 x, y , 二元函数

$$F(x, y) = P\{(X \leq x) \cap (Y \leq y)\} \stackrel{\text{记成}}{=} P\{X \leq x, Y \leq y\}$$

称为二维随机变量 (X, Y) 的分布函数, 或称为随机变量 X 和 Y 的 **联合分布函数**.

性质: 1. $F(x, y)$ 是变量 x 和 y 的不减函数, 即对于固定的 y , 当 $x_2 > x_1$ 时, $F(x_2, y) \geq F(x_1, y)$; 对于任意固定的 x , 当 $y_2 > y_1$ 时, $F(x, y_2) \geq F(x, y_1)$.

2. $0 \leq F(x, y) \leq 1$, 且

对于任意固定的 y , $F(-\infty, y) = 0$,

对于任意固定的 $y, F(x, -\infty) = 0,$

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

3. $F(x+0, y) = F(x, y), F(x, y+0) = F(x, y)$, 即 $F(x, y)$ 关于 x 右连续, 关于 y 也右连续.

4. 对于任意 $(x_1, y_1), (x_2, y_2), x_1 < x_2, y_1 < y_2$, 下述不等式成立:

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

如果二维随机变量 (X, Y) 全部可能取到的值是有限对或可列无限多对, 则称 (X, Y) 是 **离散型的随机变量**.

我们称 $P\{X = x_i, Y = y_j\} = p_{ij}, i, j = 1, 2, \dots$ 为二位离散型随机变量 (X, Y) 的 **分布律**, 或称随机变量 X 和 Y 的 **联合分布律**.

对于二维随机变量 (X, Y) 的分布函数 $F(x, y)$, 如果存在非负可积函数 $f(x, y)$, 使对于任意 x, y 有

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv,$$

则称 (X, y) 是 **连续型的二维随机变量**, 函数 $f(x, y)$ 称为二维随机变量 (X, Y) 的 **概率密度**, 或称为随机变量 X 和 Y 的 **联合概率密度**.

性质: 1. $f(x, y) \geq 0$

$$2. \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy = F(\infty, \infty) = 1.$$

3. 设 G 是 xOy 平面上的区域, 点 (X, Y) 落在 G 内的概率为

$$P\{(X, Y) \in G\} = \iint_G f(x, y) dx dy.$$

4. 若 $f(x, y)$ 在点 (x, y) 连续, 则有

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

设 E 是一个随机试验, 它的样本空间是 $S = \{e\}$, 设 $X_1 = X_1(e), X_2 = X_2(e), \dots, X_n = X_n(e)$ 是定义在 S 上的随机变量, 由它们构成的一个 n 维向量 (X_1, X_2, \dots, X_n) 叫做 **n 维随即向量**或 **n 维随机变量**.

对于任意 n 个实验 x_1, x_2, \dots, x_n, n 元函数

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

称为 n 维随机变量 (X_1, X_2, \dots, X_n) 的 **分布函数**, 或随机变量 X_1, X_2, \dots, X_n 的 **联合分布函数**.

3.2 边缘分布

二维随机变量 (X, Y) 作为一个整体, 具有分布函数 $F(x, y)$. 而 X 和 Y 都是随机变量, 各自也有分布函数, 将它们分别记为 $F_X(x), F_Y(y)$, 依次称为二维随机变量 (X, Y) 关于 X 和关于 Y 的 **边缘分布函数**.

对于离散型随机变量可得

$$F_X(x) = F(x, \infty) = \sum_{x_i \leq x} \sum_{j=1}^{\infty} p_{ij}$$

X 的分布律为

$$P\{X = x_i\} = \sum_{j=1}^{\infty} p_{ij}, \quad i = 1, 2, \dots$$

同样 Y 的分布律为

$$P\{Y = y_j\} = \sum_{i=1}^{\infty} p_{ij}, \quad j = 1, 2, \dots$$

记

$$p_{i\cdot} = \sum_{j=1}^{\infty} p_{ij} = P\{X = x_i\}, \quad i = 1, 2, \dots,$$

$$p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij} = P\{Y = y_j\}, \quad j = 1, 2, \dots,$$

分别称 $p_{i\cdot} (i = 1, 2, \dots)$ 和 $p_{\cdot j} (j = 1, 2, \dots)$ 为 (X, Y) 关于 X 和关于 Y 的 **边缘分布律**.

对于连续型随机变量 (X, Y) 设它的概率密度为 $f(x, y)$, 由于

$$F_X(x) = F(x, \infty) = \int_{-\infty}^x \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx,$$

X 是一个连续型随机变量, 其概率密度为

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

同样, Y 也是一个连续型随机变量, 其概率密度为

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

分别称 $f_X(x)$, $f_Y(y)$ 为 (X, Y) 关于 X 和关于 Y 的 **边缘概率密度**.

3.3 条件分布

设 (X, Y) 是二位离散型随机变量, 对于固定的 j , 若 $P\{Y = y_j\} > 0$, 则称

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

为在 $Y = y_j$ 条件下随机变量 X 的 **条件分布律**.

同样, 对于固定的 i , 若 $P\{X = x_i\} > 0$, 则称

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

为在 $X = x_i$ 条件下随机变量 Y 的 **条件分布律**.

设二维随机变量 (X, Y) 的概率密度为 $f(x, y)$, (X, Y) 关于 Y 的边缘概率密度为 $f_Y(y)$. 若对于固定的 y , $f_Y(y) > 0$, 则称 $\frac{f(x, y)}{f_Y(y)}$ 为在 $Y = y$ 的条件下 X 的 **条件概率密度**, 记为

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

类似地, 可以定义

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

3.4 相互独立的随机变量

设 $F(x, y)$ 及 $F_X(x), F_Y(y)$ 分别是二维随机变量 (X, Y) 的分布函数及边缘分布函数. 若对于所有 x, y 有

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\},$$

即

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

则称随机变量 X 和 Y 是相互独立的.

定理: 设 (X_1, X_2, \dots, X_n) 和 (Y_1, Y_2, \dots, Y_n) 相互独立, 则 $X_i (i = 1, 2, \dots, m)$ 和 $Y_j (j = 1, 2, \dots, n)$ 相互独立. 又若 h, g 是连续函数, 则 $g(X_1, X_2, \dots, X_m)$ 和 $g(Y_1, Y_2, \dots, Y_n)$ 相互独立.

4 第四章 随机变量的数字特征

4.1 数学期望

设离散型随机变量 X 的分布律为

$$P\{X = x_k\} = p_k \quad k = 1, 2, \dots$$

若级数

$$\sum_{k=1}^{\infty} x_k p_k$$

绝对收敛, 则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 的和为随机变量 X 的 **数学期望**, 记为 $E(X)$. 即

$$E(X) = \sum_{k=1}^{\infty} x_k p_k.$$

设连续型随机变量 X 的概率密度为 $f(x)$, 若积分

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

绝对收敛, 则称积分 $\int_{-\infty}^{\infty} x f(x) dx$ 的值为随机变量 X 的 **数学期望**, 记为 $E(X)$. 即

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

数学期望简称 **期望**, 又称 **均值**.

性质: 1. 设 C 是常数, 则有 $E(C) = C$.

2. 设 X 是随机变量, C 是常数, 则有

$$E(CX) = CE(X).$$

3. 设 X, Y 是两个随机变量, 则有

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

这一性质可以推广到任意有限多个随机变量之和的情况.

4. 设 X, Y 是相互独立的随机变量, 则有

$$E(XY) = E(X)E(Y).$$

这一性质可以推广到任意有限个相互独立的随机变量之积的情况.

4.2 方差

设 X 是一个随机变量, 若 $E\{[X - E(X)]^2\}$ 存在, 则称 $E\{[X - E(X)]^2\}$ 为 X 的 **方差**, 记为 $D(X)$ 或 $Var(X)$, 即

$$D(X) = Var(X) = E\{[X - E(X)]^2\}.$$

在应用上还引入量 $\sqrt{D(X)}$, 记为 $\sigma(X)$, 称为 **标准差**或 **均方差**.

性质: 1. 设 C 是常数, 则有 $E(C) = 0$.

2. 设 X 是随机变量, C 是常数, 则有

$$D(CX) = C^2 D(X),$$

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

3. 设 X, Y 是两个随机变量, 则有

$$D(X + Y) = D(X) + D(Y) + 2E\{(X - E(X))(Y - E(Y))\}.$$

特别, 若 X, Y 相互独立, 则有

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

这一性质可以推广到任意有限多个相互独立的随机变量之和的情况.

4. $D(X) = 0$ 的充要条件是 X 以概率为 1 取常数 $E(X)$, 即

$$P\{X = E(X)\} = 1.$$

4.3 协方差及相关系数

量 $E\{[X - E(X)][Y - E(Y)]\}$ 称为随机变量 X 与 Y 的 **协方差**, 记为 $Cov(X, Y)$, 即

$$Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\}.$$

而

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为随机变量 X 与 Y 的 **相关系数**.

由定义知

$$Cov(X, Y) = Cov(Y, X),$$

$$\text{Cov}(X, X) = D(X).$$

对于任意两个随机变量 X 和 Y ，下列等式成立：

$$D(X + Y) = D(X) + D(Y) + 2\text{Cov}(X, Y).$$

将 $\text{Cov}(X, Y)$ 的定义式展开，易得

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

我们常常利用这一式子计算协方差.

性质: 1. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$, a, b 是常数.

2. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.

定理: 1. $|\rho_{XY}| \leq 1$.

2. $|\rho_{XY}| = 1$ 的充要条件是，存在常数 a, b 使

$$P\{Y = a + bX\} = 1.$$

当 $|\rho_{XY}| = 1$ 时，称 X 和 Y **不相关**.

4.4 矩、协方差矩阵

设 X 和 Y 是随机变量，若

$$E(X^k), \quad k = 1, 2, \dots$$

存在，称它为 X 的 k **阶原点矩**，简称 k 阶矩. 若

$$E\{[X - E(X)]^k\}, \quad k = 2, 3, \dots$$

存在，称它为 X 的 k **阶中心矩**. 若

$$E(X^k Y^l), \quad k, l = 1, 2, \dots$$

存在，称它为 X 和 Y 的 $k + l$ **阶混合矩**. 若

$$E\{[X - E(X)]^k [Y - E(Y)]^l\}, \quad k, l = 1, 2, \dots$$

存在，称它为 X 和 Y 的 $k + l$ **阶混合中心矩**.

显然， X 的数学期望 $E(X)$ 是 X 的一阶原点矩，方差 $D(X)$ 是 X 的二阶中心矩，协方差 $\text{Cov}(X, Y)$ 是 X 和 Y 的二阶混合中心矩.

设 n 维随机变量 (X_1, X_2, \dots, X_n) 的二阶混合中心矩

$$c_{ij} = \text{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}, i, j = 1, 2, \dots, n$$

都存在, 则称矩阵

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为 n 维随机变量 (X_1, X_2, \dots, X_n) 的 **协方差矩阵**. 由于 $c_{ij} = c_{ji}$ ($i \neq j; i, j = 1, 2, \dots, n$), 因而上述矩阵是一个对阵矩阵.

5 第五章 大数定律及中心极限定理

5.1 大数定律

弱大数定理 (辛钦大数定理): 设 X_1, X_2, \dots 是相互独立, 服从同一分布的随机变量序列, 且具有数学期望 $E(X_k) = \mu$ ($k = 1, 2, \dots$). 作前 n 个变量的算术平均 $\frac{1}{n} \sum_{k=1}^n X_k$. 则对于任意 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| < \epsilon\right\} = 1.$$

设 $Y_1, Y_2, \dots, Y_n, \dots$ 是一个随机变量序列, a 是一个常数. 若对于任意正数 ϵ , 有

$$\lim_{n \rightarrow \infty} P\{|Y_n - a| < \epsilon\} = 1,$$

则称序列 $Y_1, Y_2, \dots, Y_n, \dots$ **依概率收敛于** a , 记为

$$Y_n \xrightarrow{P} a.$$

性质: 设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, 又设函数 $g(x, y)$ 在点 (a, b) 连续, 则

$$g(X_n, Y_n) \xrightarrow{P} g(a, b).$$

弱大数定理 (辛钦大数定理) 又可叙述为: 设随机变量 X_1, X_2, \dots 相互独立, 服从同一分布, 且具有数学期望 $E(X_k) = \mu$ ($k = 1, 2, \dots$). 则序列 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ 依概率收敛于 μ , 即 $\bar{X} \xrightarrow{P} \mu$.

伯努利大数定理: 设 f_A 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 在每次试验中发生的概率, 则对任意正数 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\{|\frac{f_A}{n} - p| < \epsilon\} = 1$$

或

$$\lim_{n \rightarrow \infty} P\{|\frac{f_A}{n} - p| \geq \epsilon\} = 0.$$

5.2 中心极限定理

独立同分布的中心极限定理: 设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 服从同一分布, 且具有数学期望和方差 $E(X_k) = \mu, D(X_k) = \sigma^2 > 0 (k = 1, 2, \dots)$, 则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

$$Y_n = \frac{\sum_{k=1}^n X_k - E(\sum_{k=1}^n X_k)}{\sqrt{D(\sum_{j=1}^n X_k)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

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

棣莫弗-拉普拉斯 (De Moivre-Laplace) 定理: 设随机变量 $\eta_n (n = 1, 2, \dots)$ 服从参数为 $n, p (0 < p < 1)$ 的二项分布, 则对于任意 x , 有

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

6 第六章 样本及抽样分布

6.1 随机样本

试验的全部可能的观察值称为 **总体**, 每一个可能观察值称为 **个体**, 总体中所包含的个体的个数称为总体的 **容量**. 容量为有限的称为 **有限总体**, 容量为无限的称为 **无限总体**.

设 X 是具有分布函数 F 的随机变量, 若 X_1, X_2, \dots, X_n 是具有同一分布函数 F 的、相互独立的随机变量 X_1, X_2, \dots, X_n 为从分布函数 F (或总体 F 、或总体 X) 得到的 **容量为 n 的简单随机样本**, 简称 **样本**, 它们的观察值 x_1, x_2, \dots, x_n 称为 **样本值**, 又称为 X 的 n 个独立的观察值.

6.2 抽样分布

设 X_1, X_2, \dots, X_n 是来自总体 X 的一个样本, $g(X_1, X_2, \dots, X_n)$ 是 X_1, X_2, \dots, X_n 的函数, 若 g 中不含未知参数, 则称 $g(X_1, X_2, \dots, X_n)$ 是 **统计量**.

因为 X_1, X_2, \dots, X_n 都是随机变量, 而统计量 $g(X_1, X_2, \dots, X_n)$ 是随机变量的函数, 因此统计量也是一个随机变量. 设 x_1, x_2, \dots, x_n 是相应于样本 X_1, X_2, \dots, X_n 的样本值, 则称 $g(x_1, x_2, \dots, x_n)$ 是 $g(X_1, X_2, \dots, X_n)$ 的观察值.

设 X_1, X_2, \dots, X_n 是来自总体 X 的一个样本, x_1, x_2, \dots, x_n 是这一样本的观察值. 定义

样本 (平) 均值:

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

样本方差:

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

样本标准差:

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

样本 k 阶 (原点) 矩:

$$A_k = \frac{1}{n} \sum_{i=1}^k, \quad k = 1, 2, \dots;$$

样本 k 阶中心矩:

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k = 2, 3, \dots.$$

样本经验分布函数又称样本分布函数, 记为 $F_n(x)$, 定义为

$$F_n(x) = \frac{1}{n} (\#X_i \leq x), \quad -\infty < x < \infty,$$

其中 $(\#X_i \leq x)$ 表示 X_1, X_2, \dots, X_n 中小于或等于 x 的个数.

格里汶科定理: 设 X_1, X_2, \dots, X_n 是来自以 $F(x)$ 为分布函数的总体 X 的样本, $F_n(x)$ 是样本经验分布函数, 则有

$$P\left\{\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = 0\right\} = 1.$$

统计量的分布称为 **抽样分布**.

6.2.1 χ^2 分布

设 X_1, X_2, \dots, X_n 是来自总体 $N(0, 1)$ 的样本, 则称统计量

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

服从自由度为 n 的 χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$.

χ^2 分布的概率密度为

$$f(y) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, & y > 0, \\ 0, & \text{其他.} \end{cases}$$

χ^2 分布的可加性: 设 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 并且 χ_1^2, χ_2^2 相互独立, 且有

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2).$$

χ^2 分布的数学期望和方差: 若 $\chi^2 \sim \chi^2(n)$, 则有

$$E(\chi^2) = n, \quad D(\chi^2) = 2n.$$

χ^2 分布的上分位点: 对于给定正数 $\alpha, 0 < \alpha < 1$, 称满足条件

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^{\infty} f(y) dy = \alpha$$

的点 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 分布的上 α 分位点.

6.2.2 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

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

服从自由度为 n 的 t 分布. 记为 $t \sim t(n)$.

t 分布又称 **学生氏 (Student) 分布**. $t(n)$ 分布的概率密度为

$$h(t) = \frac{\Gamma[\frac{n+1}{2}]}{\sqrt{\pi n} \Gamma(\frac{n}{2})} (1 + \frac{t^2}{n})^{-\frac{n+1}{2}}, \quad -\infty < t < \infty$$

t 分布的上分位点: 对于给定的 $\alpha, 0 < \alpha < 1$, 称满足条件

$$P\{t > t_\alpha(n)\} = \int_{t_\alpha(n)}^{\infty} h(t) dt = \alpha$$

的点 $t_\alpha(n)$ 为 $t(n)$ 的上 α 分位点.

6.2.3 F 分布

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$, 且 U, V 相互独立, 则称随机变量

$$F = \frac{\frac{U}{n_1}}{\frac{V}{n_2}}$$

服从自由度为 (n_1, n_2) 的 F 分布, 记为 $F \sim F(n_1, n_2)$.

$F(n_1, n_2)$ 分布的概率密度为

$$\psi(y) = \begin{cases} \frac{\Gamma(\frac{n_1+n_2}{2}) (\frac{n_1}{n_2})^{\frac{n_1}{2}} y^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2}) (1 + \frac{n_1 y}{n_2})^{\frac{n_1+n_2}{2}}}, & y > 0, \\ 0, & \text{其他.} \end{cases}$$

F 分布的上分位点对于给定的 $\alpha, 0 < \alpha < 1$, 称满足条件

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^{\infty} \psi(y) dy = \alpha$$

的点 $F_\alpha(n_1, n_2)$ 为 $F(n_1, n_2)$ 分布的上 α 分位点.

6.2.4 正态总体的样本均值与样本方差的分布

设总体 X (不管服从什么分布, 只要均值和方差存在) 的均值为 μ , 方差为 $\sigma^2, X_1, X_2, \dots, X_n$ 是来自 X 的一个样本, \bar{X}, S^2 分别是样本均值和样本方差, 则有

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

而

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

即

$$E(S^2) = \sigma^2.$$

进而, 设 $X \sim N(\mu, \sigma^2)$, 知 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 也服从正态分布, 于是得到以下的定理:

定理一: 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, \bar{X} 是样本均值, 则有

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

定理二: 设 X_1, X_2, \dots, X_n 是来自总体 $N(\mu, \sigma^2)$ 的样本, \bar{X}, S^2 分别是样本均值和样本方差, 则有

1. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$;
2. \bar{X} 与 S^2 相互独立.

定理三: 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, \bar{X} 是样本均值, 则有

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

定理四: 设 X_1, X_2, \dots, X_{n_1} 与 Y_1, Y_2, \dots, Y_{n_2} 分别是来自正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的样本, 且这两个样本相互独立. 设 $S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$, $S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$ 分别是这两个样本的样本方差, 则有

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

7 第七章 参数估计

7.1 点估计

7.1.1 矩估计法

设 X 为连续型随机变量, 其概率密度为 $f(x; \theta_1, \theta_2, \dots, \theta_k)$, 或 X 为离散型随机变量, 其分布律为 $P\{X = x\} = p(x; \theta_1, \theta_2, \dots, \theta_k)$, 其中 $\theta_1, \theta_2, \dots, \theta_k$ 为待估参数, X_1, X_2, \dots, X_n 是来自 X 的样本. 假设总体 X 的前 k 阶矩

$$\mu_l = E(X^l) = \int_{-\infty}^{\infty} x^l f(x; \theta_1, \theta_2, \dots, \theta_k) dx \quad (X \text{ 连续型})$$

或

$$\mu_l = E(X^l) = \sum_{x \in R_X} x^l p(x; \theta_1, \theta_2, \dots, \theta_k) \quad (X \text{ 离散型}), \quad l = 1, 2, \dots, k$$

(其中 R_X 是 X 可能取值的范围) 存在. 一般来说他们是 $\theta_1, \theta_2, \dots, \theta_k$ 的函数. 基于样本矩

$$A_l = \frac{1}{n} \sum_{i=1}^n X_i^l$$

依概率收敛于相应的总体矩 $\mu_l = (l = 1, 2, \dots, k)$, 样本矩的连续函数一概率收敛于相应的总体矩的连续函数, 我们就用样本矩作为相应的总体矩的估计量, 而以样本矩的连续函数作为相应的总体矩的连续函数的估计量. 这种估计方法称为 **矩估计法**. 矩估计法的具体做法如下: 设

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \dots, \theta_k), \\ \mu_2 = \mu_2(\theta_1, \theta_2, \dots, \theta_k), \\ \vdots \\ \mu_k = \mu_k(\theta_1, \theta_2, \dots, \theta_k). \end{cases}$$

这是一个包含 k 个未知参数 $\theta_1, \theta_2, \dots, \theta_k$ 的联立方程组. 一般来说, 可以从中间解出 $\theta_1, \theta_2, \dots, \theta_k$, 得到

$$\begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \dots, \mu_k), \\ \theta_2 = \theta_2(\mu_1, \mu_2, \dots, \mu_k), \\ \vdots \\ \theta_k = \theta_k(\mu_1, \mu_2, \dots, \mu_k). \end{cases}$$

以 A_i 分别代替上式中的 $\mu_i, i = 1, 2, \dots, k$, 就以

$$\hat{\theta}_i = \theta_i(A_1, A_2, \dots, A_k), \quad i = 1, 2, \dots, k$$

分别作为 $\theta_i, i = 1, 2, \dots, k$ 的估计量, 这种估计量称为 **矩估计量**. 矩估计量的观察值称为 **矩估计值**.

7.1.2 最大似然估计法

若总体 X 属离散型, 其分布律 $P\{X = x\} = p(x; \theta), \theta \in \Theta$ 的形式为已知, θ 为待估参数, Θ 是 θ 可能取值的范围. 设 X_1, X_2, \dots, X_n 是来自 X 的样本,

则 X_1, X_2, \dots, X_n 的联合分布律为

$$\prod_{i=1}^n p(x_i; \theta).$$

又设 x_1, x_2, \dots, x_n 是相应于样本 X_1, X_2, \dots, X_n 的一个样本值. 易知样本 X_1, X_2, \dots, X_n 取到的观察值 x_1, x_2, \dots, x_n 的概率, 亦即事件 $\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 发生的概率为

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta), \theta \in \Theta.$$

这一概率随 θ 的取值而变化, 它是 θ 的函数, $L(\theta)$ 称为样本的 **似然函数** (注意, 这里 x_1, x_2, \dots, x_n 是已知的样本值, 他们都是常数). 由费希尔引进的最大似然估计法, 就是固定样本观察值 x_1, x_2, \dots, x_n 在 θ 取值的可能范围 Θ 内挑选使似然函数 $L(x_1, x_2, \dots, x_n; \theta)$ 达到最大的参数值 $\hat{\theta}$, 作为参数 θ 的估计值. 即取 $\hat{\theta}$ 使

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta).$$

这样得到的 $\hat{\theta}$ 与样本值 x_1, x_2, \dots, x_n 有关, 常记为 $\hat{\theta}(x_1, x_2, \dots, x_n)$, 称为参数 θ 的 **最大似然估计值**. 而相应的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 称为参数 θ 的 **最大似然估计量**.

若总体 X 属连续型, 其概率密度 $f(x; \theta), \theta \in \Theta$ 的形式已知, θ 为待估参数, Θ 是 θ 可能取值的范围. 设 X_1, X_2, \dots, X_n 是来自 X 的样本, 则 X_1, X_2, \dots, X_n 的联合密度为

$$\prod_{i=1}^n f(x_i, \theta).$$

设 x_1, x_2, \dots, x_n 是相应于样本 X_1, X_2, \dots, X_n 的一个样本值, 则随机点 (X_1, X_2, \dots, X_n) 落在点 (x_1, x_2, \dots, x_n) 的邻域 (边长分别为 dx_1, dx_2, \dots, dx_n 的 n 维立方体) 内的概率近似地为

$$\prod_{i=1}^n f(x_i; \theta) dx_i.$$

其值随 θ 的取值而变化. 与离散型的情况一样, 我们取 θ 的估计值 $\hat{\theta}$ 使概率取到最大值, 但因因子 $\prod_{i=1}^n dx_i$ 不随 θ 而变, 故只需考虑函数

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

的最大值. 这里 $L(\theta)$ 称为样本的 **似然函数**. 若

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta),$$

则称 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为 θ 的 **最大似然函数估计值**, 称 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为 θ 的 **最大似然估计值**. 这样, 确定最大似然估计量的问题就归结为微分学中的求最大值的问题了. 在很多情况下, $p(x; \theta)$ 和 $f(x; \theta)$ 关于 θ 可微, 这是 $\hat{\theta}$ 常可从方程

$$\frac{d}{d\theta} L(\theta) = 0$$

解得. 又因 $L(\theta)$ 与 $\ln L(\theta)$ 在同一 θ 处取到极值, 因此, θ 的最大似然估计 θ 也可以从 **对数似然方程**

$$\frac{d}{d\theta} \ln L(\theta) = 0$$

求得.

7.2 估计量的评选标准

7.2.1 无偏性

无偏性: 若估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在, 且对于任意 $\theta \in \Theta$ 有

$$E(\hat{\theta}) = \theta,$$

则称 $\hat{\theta}$ 是 θ 的 **无偏估计量**.

7.2.2 有效性

有效性: 设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \dots, X_n)$ 与 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \dots, X_n)$ 都是 θ 的无偏估计量, 若对于任意 $\theta \in \Theta$, 有

$$D(\hat{\theta}_1) \leq D(\hat{\theta}_2),$$

且至少对于某一个 $\theta \in \Theta$ 上述不等式成立, 则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ **有效**.

7.2.3 相合性

相合性: 设 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为参数 θ 的估计量, 若对于任意 $\theta \in \Theta$, 当 $n \rightarrow \infty$ 时, $\hat{\theta}(X_1, X_2, \dots, X_n)$ 依概率收敛于 θ , 则称 $\hat{\theta}$ 为 θ 的 **相合估计量**. 即若对于任意 $\theta \in \Theta$ 都满足: 对于任意 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta} - \theta| < \epsilon\} = 1,$$

则称 $\hat{\theta}$ 是 θ 的 **相合估计量**.

7.3 区间估计

设总体 X 的分布函数 $F(x; \theta)$ 含有一个未知参数 $\theta, \theta \in \Theta$ (Θ 是 θ 可能取值的范围), 对于给定值 $\alpha (0 < \alpha < 1)$, 若由来自 X 的样本 X_1, X_2, \dots, X_n 确定的两个统计量 $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$ 和 $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n) (\underline{\theta} < \bar{\theta})$, 对于任意 $\theta \in \Theta$ 满足

$$P\{\underline{\theta}(X_1, X_2, \dots, X_n) < \theta < \bar{\theta}(X_1, X_2, \dots, X_n)\} \geq 1 - \alpha,$$

则称随机区间 $(\underline{\theta}, \bar{\theta})$ 是 θ 的置信水平为 $1 - \alpha$ 的 **置信区间**, $\underline{\theta}$ 和 $\bar{\theta}$ 分别称为置信水平为 $1 - \alpha$ 的双侧置信区间的 **置信下限**和 **置信上限**, $1 - \alpha$ 称为 **置信水平**.

求未知参数 θ 的置信区间的具体做法如下.

1. 寻求一个样本 X_1, X_2, \dots, X_n 和 θ 的函数 $W = W(X_1, X_2, \dots, X_n; \theta)$, 使得 W 的分布不依赖于 θ 以及其他未知参数, 称具有这种性质的函数 W 为 **枢轴量**.
2. 对于给定的置信水平 $1 - \alpha$ 给定两个常数 a, b 使得

$$P\{a < W(X_1, X_2, \dots, X_n; \theta) < b\} = 1 - \alpha.$$

若能从 $a < W(X_1, X_2, \dots, X_n; \theta) < b$ 得到与之等价的 θ 的不等式 $\underline{\theta} < \theta < \bar{\theta}$, 其中 $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n), \bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$ 都是统计量. 那么 $(\underline{\theta}, \bar{\theta})$ 就是 θ 的一个置信水平为 $1 - \alpha$ 的置信区间.

枢轴量 $W(X_1, X_2, \dots, X_n; \theta)$ 的构造, 通常可以从 θ 的点估计着手考虑. 通常的正态总体的参数的置信区间可以用上述步骤推得.

7.4 正态总体均值与方差的区间估计

7.4.1 单个总体 $N(\mu, \sigma^2)$ 的均值 μ 的置信区间

1. σ^2 为已知

采用枢轴量 $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$, 得 μ 的一个置信水平为 $1 - \alpha$ 的置信区间为

$$(\bar{X} - \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}}, \bar{X} + \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}});$$

2. σ^2 为未知

采用枢轴量 $\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$, 得 μ 的一个置信水平为 $1 - \alpha$ 的置信区间为

$$(\bar{X} - \frac{St_{\frac{\alpha}{2}}(n-1)}{\sqrt{n}}, \bar{X} + \frac{St_{\frac{\alpha}{2}}(n-1)}{\sqrt{n}}).$$

7.4.2 两个总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ 的均值差 $\mu_1 - \mu_2$ 的置信区间

1. σ_1^2, σ_2^2 均为已知
2. σ_1^2, σ_2^2 均未知