

概率论笔记

一个短篇

2024 年 01 月 16 日

目录

| | |
|-------------------------------|-----------|
| 1 基本概念 | 2 |
| 1.1 运算 | 2 |
| 1.2 关系 | 2 |
| 1.3 频率和概率 | 2 |
| 1.4 条件概率 | 3 |
| 1.5 全概率与贝叶斯公式 | 3 |
| 1.6 事件的独立性 | 3 |
| 1.7 伯努利概型 | 3 |
| 2 随机变量及其分布 | 4 |
| 2.1 $R.V.$ 和 分布函数 | 4 |
| 2.2 分布函数 | 4 |
| 2.3 离散型随机变量及其分布 | 4 |
| 2.4 连续型随机变量 | 5 |
| 2.5 随机变量的分布 | 6 |
| 3 多维 $R.V.$ | 6 |
| 3.1 二维随机变量和联合分布函数 | 6 |
| 3.2 连续型随机变量 | 7 |
| 3.3 边缘分布 | 7 |
| 3.4 条件分布 | 9 |
| 3.5 独立性 | 9 |
| 3.6 二维连续型随机变量的分布 | 9 |
| 4 随机变量的数字特征 | 10 |
| 4.1 数学期望 | 10 |
| 4.2 方差 | 10 |
| 4.3 常见形式 | 11 |
| 4.4 协方差 | 11 |
| 5 大数定律和中心极限定理 | 12 |
| 5.1 大数定律 | 12 |
| 5.2 中心极限定理 | 13 |
| 6 样本与抽样分布 | 13 |
| 6.1 基本概念 | 13 |
| 6.2 抽样分布 | 14 |
| 7 参数估计 | 15 |
| 7.1 点估计 | 15 |
| 7.2 评选标准 | 15 |
| 7.3 区间估计 | 15 |

1 基本概念

1.1 运算

若 A 代表事件 A 发生, \bar{A} 代表事件没有发生, 我们定义如下在随机事件上的关系运算:

包含

$$A \subset B \Leftrightarrow A \rightarrow B$$

$$\text{另外有 } A = B \Leftrightarrow A \subset B \wedge B \subset A$$

差集

$$A - B \Leftrightarrow A \wedge \bar{B}$$

$$\text{另外有 } A - B = A - AB = A\bar{B}$$

交集

$$A \cap B \Leftrightarrow A \wedge B$$

并集

$$A \cup B \Leftrightarrow A \vee B$$

对于交集和并集运算, 符合以下四种运算律:

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

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

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

对偶律 (德摩根定律)

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

1.2 关系

在事件间, 存在如下两种关系:

互斥事件

$$A \cap B = \emptyset$$

对立事件

$$A \cap B = \emptyset \wedge A \cup B = \Omega.$$

1.3 频率和概率

Definition 1.1:

频率

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

概率

$$P(A) = \lim_{n \rightarrow \infty} f_n(A) \rightarrow P$$

Properties 1.1:

非负性

$$0 \leq P(A) \leq 1$$

规范性

$$P(\Omega) = 1, P(\emptyset) = 0$$

无限可加性

$$\text{If } A \cap B = \emptyset$$

$$\text{then } P(A \cup B) = P(A) + P(B)$$

互补性

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

概率之间存在如下运算：

Definition 1.2（运算）：

1. 减法： $P(A - B) = P(A) - P(AB)$, If $B \subset A$, $P(A - B) = P(A) - P(B)$
2. 加法： $P(A + B) = P(A \cup B) = P(A) + P(B) - P(AB)$
3. 乘法： $P(AB) = P(A \cap B) = P(A)P(B | A)$

1.4 条件概率

Definition 1.3：如果 $P(A) > 0$ ，则条件概率为 $P(B | A) = \frac{P(AB)}{P(A)}$ 。

依此，我们有两条推广式：

1. $P(B \cup C | A) = P(B | A) + P(C | A) - P(BC | A)$
2. $P(B - C | A) = P(B | A) - P(BC | A)$

1.5 全概率与贝叶斯公式

Definition 1.4（完备事件组）： $S = \{A_1, A_2, A_3, \dots, A_n\}$ 是一个属于 Ω 的事件组，并且满足 $\forall A_1, A_2 \subset S, A_1 \cap A_2 = \emptyset, \wedge A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n = \Omega$ ，则 S 为一个完备事件组。

由定义，我们可以设 B 是一个随机事件， $\{A_1, A_2, \dots, A_i\}$ 是一个完备事件组，我们有：

Formula 1.1（全概率公式）：

$$\begin{aligned} P(B) &= \sum_{i=1}^n P(A_i)P(B | A_i) \\ &= \sum_{i=1}^n P(BA_i) \end{aligned}$$

Formula 1.2（贝叶斯公式）：

$$P(A_i | B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B | A_i)}{\sum_{i=1}^n P(A_i)P(B | A_i)}$$

1.6 事件的独立性

若 A, B 是相互独立事件，则有

$$\begin{aligned} P(A | B) &= P(A) \Leftrightarrow P(AB) = P(A)P(B) \\ &\Leftrightarrow P(B | A) = P(B | \overline{A}) \\ &\Leftrightarrow P(B | \overline{A}) = P(B) \end{aligned}$$

且 $A, \overline{A}, B, \overline{B}$ 也相互独立，此外有

若 A, B, C 互为独立事件 $\rightarrow A, B, C$ 两两独立

若 A, B, C 互为独立事件 \rightarrow 关于 A, B 的加法，乘法，减法

以及逆运算也分别独立与 C 和 \overline{C}

1.7 伯努利概型

Definition 1.5（伯努利实验）：实验只有两种可能结果 A, \overline{A} 的实验叫做伯努利实验。

Formula 1.3 (二项概率公式) :

$$\binom{n}{k} p^k (1-p)^{n-k}$$

2 随机变量及其分布

2.1 R.V. 和 分布函数

R.V. 是一个从随机试验 E 的样本空间 Ω 到 \mathbb{R} 的一个映射。

2.2 分布函数

Definition 2.1: 设 X 是一个随机变量, r 是任意实数, 则称事件 $\{X \leq r\}$ 的概率为 R.V. X 的分布函数, 计作 $F(r)$ 。

分布函数有如下性质:

Properties 2.1 (范围概率): $P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1)$

Properties 2.2 (增减性): $F(x)$ 是一个不减函数

Properties 2.3 (规范性): $F(-\infty) = 0, F(+\infty) = 1$

2.3 离散型随机变量及其分布

2.3.1 常见的分布

Definition 2.2 (0-1 分布): 若随机变量 X 分布服从下方分布列, 其中 $0 < p < 1$, 则称为 0-1 分布或两点分布。

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

Definition 2.3 (二项分布): 计作 $X \sim B(n, p)$

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}, k \in \{x \mid x \in \mathbb{N}^+ \cap [0, n]\}$$

Definition 2.4 (泊松分布): 计作 $X \sim P(\lambda)$

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

Definition 2.5 (几何分布): $P\{X = k\} = (1-p)^{k-1} p$

Definition 2.6 (超几何分布): 计作 $X \sim H(N, M, n)$

$$P\{X = k\} = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, k \in \{1, 2, 3, \dots, \min(M, N)\}$$

Theorem 2.1 (泊松定理) : 当 $X \sim B(n, p)$ 且 n 充分大, p 充分小时

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = np$$

2.4 连续型随机变量

Definition 2.7 (分布函数) : 若 $f(t)$ 是概率密度函数, 则分布函数 $F(x)$ 为

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

Formula 2.1 (区间概率公式) : $F\{a < X < b\} = \int_a^b f(x) dx$

2.4.1 常见形式

Definition 2.8 (均匀分布) :

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

Definition 2.9 (指数分布) :

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

指数分布具有无记忆性, 即 $P\{x > t + T \mid x > t\} = P\{x > T\}$ 。

2.4.2 正态分布

Definition 2.10 (正态分布) : 计作 $X \sim N(\mu, \sigma^2), \sigma > 0$

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

Properties 2.4 (可加性) : 若 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 则有

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Annotation 2.1: $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{A}} dx = \sqrt{A\pi}, A > 0$

Proof: 设 $X \sim N(0, \frac{A}{2})$, 因为概率分布函数具有规范性 $F(+\infty) = 1$ 即 $\int_{-\infty}^{+\infty} f(x) = 1$. 带入得

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{A\pi}} e^{-\frac{x^2}{A}} dx &= 1 \\ \frac{1}{\sqrt{A\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{A}} dx &= 1 \\ \int_{-\infty}^{+\infty} e^{-\frac{x^2}{A}} dx &= \sqrt{A\pi} \end{aligned}$$

□

Definition 2.11 (标准正态分布) : 当 $\mu = 0, \sigma^2 = 1, X \sim N(0, 1), x \in \mathbb{R}$ 时, 其为标准正态分布。

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

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

Definition 2.12 (标准化) : 若 $X \sim N(\mu, \sigma^2)$ 不满足标准正态分布, 则 $\frac{X-\mu}{\sigma} \sim N(0, 1)$

根据标准化, 如果我们想要计算一个满足非标准化的正态分布的随机变量在范围 $(a, b]$ 上的概率, 我们可以 X 先将其标准化为 $\frac{X-\mu}{\sigma}$ 并计算 $\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$ 即可。

Definition 2.13 (分位点) : μ_α 表示 $P\{x > \mu_\alpha\} = \alpha$. 并且有 $\mu_{1-\alpha} = \mu(-\alpha)$.

2.5 随机变量的分布

对于离散型随机变量, 我们可以先求出取值, 在分别对对应的取值求出概率。而对于连续性随机变量, 重点是求其密度函数: 即已知 $X \sim f_X(x), Y = g(X)$, 求 $R.V.Y$ 的分布函数 $f_Y(y)$ 。

首先介绍根据分布函数求 $R.V.Y$ 的 密度函数的方法:

Annotation 2.2 (根据分布函数法) :

1. 首先, 找到密度函数 $f_Y(y)$ 的分段点, 一般有如下两种情况
 1. $f_X(x)$ 的分段点, 带入 $g(x)$ 后得到的 y 的值, 和
 2. $y = g(x)$ 的最值
2. 其次, 根据以上分段点, 求出区间 $(l, r]$ 的 $F_Y(y)$:

$$F_Y(y) = P\{Y \leq y\} = P\{g(x) \leq y\} = \int_l^r f_X(x) dx$$

3. 最后, 对求出的分布函数求导即可得到随机变量 y 的密度函数 $f_Y(y) = F_Y'(y)$ 。

除此之外, 还可以用下方的定理中的公式来进行求解:

Formula 2.2: 设 $f_X(x)$ 随机变量 X 的密度函数, 对于随机变量 Y 有 $Y = g(X)$, 且 $g(X)$ 为单调函数, 令 $x = h(y)$ 是 $y = g(x)$ 的反函数, α, β 分别是 $g(x)$ 的最小值和最大值。则 $Y = g(X)$ 的密度函数 $f_Y(y)$ 为:

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

3 多维 $R.V.$

3.1 二维随机变量和联合分布函数

Definition 3.1 (二维随机变量) : 设随机试验 E 的样本空间 $\Omega = \{e\}$, $X = X(e), Y = Y(e)$ 的定义在 Ω 上的随机变量, 则 (X, Y) 为定义在 Ω 上的二维随机变量。

Definition 3.2 (联合分布函数)：设 $x, y \in \mathbb{R}$ ，则 x, y 的联合分布函数为事件 $\{X \leq x\}$ 与事件 $\{Y \leq y\}$ 同时发生的概率为二维随机变量的联合分布函数

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

随机变量的分布函数的几何意义为在二维坐标轴中， $F(x, y)$ 为 $X \leq x$ 与 $Y \leq y$ 所围成的矩形区域的面积。易得，点 (X, Y) 落在 $\{(x, y) \mid x_1 < x \leq x_2, y_1 < y \leq y_2\}$ 区域的概率为 $F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$ 。

3.2 连续型随机变量

Definition 3.3 (密度函数)：联合概率密度指的是对于二维随机变量 (X, Y) ，其概率分布函数为

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

其中的非负函数 $f(x, y)$ 即为联合概率密度。

二维随机变量的联合密度函数有如下性质：

Properties 3.1:

1. 非负性：

$$f(x, y) \geq 0$$

2. 规范性：

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \, du \, dv = F(\infty, \infty) = 1$$

3. 设 G 是平面 xOy 上的闭区域，则点 (X, Y) 落在 G 区域上的概率为

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

3.2.1 常见形式

Definition 3.4 (均匀分布)：

$$f(u, v) = \begin{cases} \frac{1}{S_G}, & (u, v) \in G \\ 0, & \text{其他} \end{cases}$$

3.3 边缘分布

Definition 3.5 (分布函数)：设二维随机变量 (X, Y) 的联合分布函数为 $F(X, Y)$ ， Ω 为完备事件组，则 $F_X(x) = P\{X \leq x, \Omega\}$ ， $F_Y(y) = P\{\Omega, Y \leq y\}$ 分别为二维随机变量关于 X 或 Y 的边缘分布函数。

Definition 3.6 (分布律 / 质量函数)：已知二维随机变量 (X, Y) 的分布律为 $P\{X, Y\} = P\{X = x_i \cap Y = y_i\}$ 则关于 X 的边缘分布律为

$$P_X(x) = \sum_i P(x, y_i)$$

Definition 3.7 (密度函数)：设 (X, Y) 的密度函数为 $f(x, y)$ 则关于 X 的边缘密度函数和关于 Y 的边缘密度函数为

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

3.4 条件分布

Definition 3.8 (分布律)：有二维随机变量 (X, Y)

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P_Y\{Y = y_j\}}$$

即为随机变量 X 在 $Y = y_j$ 下的条件分布律

Definition 3.9 (密度函数)：有二维随机变量 (X, Y) 及其联合概率密度 $f(x, y)$ ，固定 $Y = y$ ，则随机变量 X 在 $Y = y$ 条件下的概率密度函数为 $f_{X|Y} = \frac{f(x, y)}{f_Y(y)}$

3.5 独立性

Definition 3.10：有二维随机变量 (X, Y) ， $F(x, y) = F_X(x)F_Y(y)$ 满足独立。对于离散型随机变量，独立性在于是否满足 $P(x, y) = P_X(x)P_Y(y)$ 。对于连续性随机变量，在于其密度函数是否满足 $f(x, y) = f_X(x)f_Y(y)$ 。

3.6 二维连续型随机变量的分布

对于离散型随机变量，依旧是先求取值，再求概率，而对于两个连续型随机变量，我们有如下方法：

Formula 3.1 (分布函数法)：有二维随机变量 (X, Y) 及其联合概率分布 $f(x, y)$ ，已知 $Z = g(X, Y)$ ，则 Z 的概率分布为

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} = P\{g(X, Y) \leq z\} = P\{(X, Y) | g(X, Y) \leq z\} \\ &= \iint_G f(x, y) dx dy \end{aligned}$$

或使用卷积公式：

Formula 3.2 (卷积公式)：若随机变量 X, Z, Y 存在 $Z = X + Y$ 关系，则

| | |
|--|--|
| X, Y 不独立 $f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx$ | X, Y 独立 $f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx$ |
|--|--|

Proof: 对于 $Z = X + Y$

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

令 $y + x = t$ 将二重积分中对 y 的积分换为对 t 的积分得

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^z f(x, t-x) dt$$

改变积分次序后得

$$\int_{-\infty}^z \left[\int_{-\infty}^{+\infty} f(x, t-x) dx \right] dt$$

要求 Z 的密度函数，对变上限积分求导可得

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx$$

若 (X, Y) 独立, 又有

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx$$

□

4 随机变量的数字特征

4.1 数学期望

Definition 4.1 (离散型): 对于离散型随机变量 X , 设 x_i 为其分布律的第 i 个取值, 相应概率为 p_i , 则其数学期望 (均值) 为:

$$E(X) = \sum x_i p_i$$

Definition 4.2 (连续型): 若连续型随机变量 X 的密度函数为 $f_X(x)$ 则它的均值为

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

若要求 $Y = g(X)$ 的均值, 则

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

若 $Z = g(X, Y)$ 且二维随机变量 (X, Y) 有联合概率密度 $f(x, y)$ 则

$$E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy$$

Properties 4.1: 有常数 C , 随机变量 X 与 Y :

1. $E(C) = C$
2. $E(C + X) = C + E(X)$
3. $E(CX) = CE(X)$
4. $E(X + Y) = E(X) + E(Y)$
5. 若 X, Y 独立, 则 $E(XY) = E(X)E(Y)$

4.2 方差

Definition 4.3: 方差为随机变量与其均值的距离的平方的均值即 $D(X) = E[(X - E(X))^2]$, 其用来表示 X 偏离其均值 $E(X)$ 的程度大小。且方差 $D(X) \geq 0$ 。

Formula 4.1 (方差计算公式): $D(X) = E(X^2) - E^2(X)$

Properties 4.2: C 为常数

1. $D(C) = 0$
2. $D(X + C) = D(X)$
3. $D(CX) = C^2 D(X)$
4. $D(X + Y) = D(X) + D(Y) - 2[E(XY) - E(X)E(Y)]$

4.3 常见形式

Theorem 4.1 (0-1 分布)：若随机变量 X 服从 0-1 分布，则

$$E(X) = p, D(X) = p(1-p)$$

Theorem 4.2 (二项分布)：若随机变量 X 服从二项分布即 $X \sim B(n, p)$ 则

$$E(X) = np, D(X) = np(1-p)$$

Proof: 若随机变量 $X \sim B(n, p)$ 则其含义为 n 重伯努利实验中成功的次数即

$X = X_1 + X_2 + \dots + X_n$ ，其中 X_i 表示第 i 次伯努利实验，每次伯努利实验独立且都有相同的 p ，即 $E(X_i) = p$ ，则

$$E(X) = E(X_1 + X_2 + \dots + X_n) = np$$

$$D(X) = D(X_1 + X_2 + \dots + X_n) = np(1-p)$$

□

Theorem 4.3 (泊松分布)：若随机变量 X 服从参数为 λ 的泊松分布即 $X \sim P(\lambda)$ 则

$$E(X) = D(X) = \lambda$$

Theorem 4.4 (均匀分布)：若 $X \sim U(a, b)$ 则

$$E(X) = \frac{a+b}{2}, D(X) = \frac{(b-a)^2}{12}$$

Theorem 4.5 (指数分布)：若 $X \sim e(\lambda)$ 则

$$E(X) = \frac{1}{\lambda}, D(X) = \frac{1}{\lambda^2}$$

Theorem 4.6 (正态分布)：若 $X \sim N(\mu, \sigma^2)$ 则

$$E(X) = \mu, D(X) = \sigma^2$$

4.4 协方差

Definition 4.4 (协方差)：有二维随机变量 (X, Y) ，称 $E[(X - E(X))(Y - E(Y))]$ 为随机变量 (X, Y) 的协方差，通常计作 $\text{cov}(X, Y)$ 即

$$\begin{aligned}\text{cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

特别的，相同变量的协方差为其方差 $\text{cov}(X, X) = D(X)$ 。

已知方差的性质 3: $D(X + Y) = D(X) + D(Y) - 2[E(XY) - E(X)E(Y)]$ ，我们将协方差的计算公式带入可得到 $D(X + Y) = D(X) + D(Y) - 2\text{cov}(X, Y)$ 。

Properties 4.3：协方差有以下性质

1. $\text{cov}(X, Y) = \text{cov}(Y, X)$

2. $\text{cov}(aX, bY) = ab \text{cov}(X, Y)$

3. $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$

4. 若 X, Y 相互独立，则 $\text{cov}(X, Y) = 0$

Definition 4.5 (相关系数)：有随机变量 X, Y ，则其相关系数为：

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

Properties 4.4：相关系数有以下性质

1. $-1 \leq \rho_{XY} \leq 1$
2. 相关性
 1. 若相关系数 $\rho_{XY} = 0$ 则称 X, Y 不相关
 2. 若 $\rho_{XY} = 1$ ，则称 X, Y 为正相关， $y = ax + b, a > 0$
 3. 若 $\rho_{XY} = -1$ ，则称 X, Y 为负相关， $y = ax + b, a < 0$

5 大数定律和中心极限定理

5.1 大数定律

Theorem 5.1 (切比雪夫不等式)：有随机变量 X 及其均值 $E(X)$ 方差 $D(X)$ ，存在任意正数 ε 有

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}$$

另有

$$P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$$

Theorem 5.2 (切比雪夫大数定律)：设 X_1, X_2, \dots, X_n 是相互独立的随机变量序列，且 $E(X_i), D(X_i)$ 均存在，且 $D(X_i) \leq C$ ，记 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 则对于任意正数 ε ，有

$$\begin{aligned} \lim_{n \rightarrow +\infty} P\{|\bar{X} - E(\bar{X})| < \varepsilon\} &= 1 \\ \Leftrightarrow \bar{X} \text{ 依概率收敛到 } E(\bar{X}) \text{ 即: } \bar{X} \xrightarrow{P} E(\bar{X}) \end{aligned}$$

Theorem 5.3 (伯努利大数定律)：设 n_A 为 n 次独立重复试验中事件 A 发生的次数，且 $P(A) = p$ ，则对于任意正数 ε ，有

$$\begin{aligned} \lim_{n \rightarrow +\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} &= 1 \\ \Leftrightarrow \frac{n_A}{n} \xrightarrow{P} p \end{aligned}$$

Theorem 5.4 (辛钦大数定律)：有随机变量序列 X_1, X_2, \dots, X_n ，随机变量间相互独立且服从同一分布， $E(X_i)$ 存在，则对于任意正数 ε 有

$$\begin{aligned} \lim_{n \rightarrow +\infty} P\{|\bar{X} - E(\bar{X})| < \varepsilon\} &= 1 \\ \Leftrightarrow \bar{X} \xrightarrow{P} E(\bar{X}) \end{aligned}$$

5.2 中心极限定理

独立随机变量的和的分布当随机变量的个数足够大的时候，近似的服从正态分布。

Theorem 5.5 (独立同分布 (列维 – 林德伯格) 中心极限定理)：若一随机变量序列 X_1, X_2, \dots, X_n 服从同一分布且具有相同的期望 $E(X_i) = \mu$ ，相同的方差 $D(X_i) = \sigma^2$ ，将 $\sum_{i=1}^n X_i$ 计作 η_n 则当 n 充分大的时候，有

$$\begin{aligned}\eta_n &\text{ 近似服从 } N(E(\eta_n), D(\eta_n)) \\ &= N(n\mu, n\sigma^2)\end{aligned}$$

又由正态分布的标准化可得

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ 近似服从 } N(0, 1)$$

Theorem 5.6 (棣莫弗 – 拉普拉斯 (De Moivre – Laplace) 定理)：若随机变量 $X_n, n = 1, 2, \dots$ 服从参数为 n, p 的二项分布，也即随机变量 X 可以分为 n 的相互独立的随机变量 X_i 服从 $0-1$ 分布，对于任意的 x 有

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

也即：当 n 充分大时， $\sum_{i=1}^n X_i$ 近似服从参数为 np 与 $np(1-p)$ 的正态分布，进而 $\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}}$ 近似服从标准正态分布。

6 样本与抽样分布

6.1 基本概念

Definition 6.1 (样本)：设随机变量 X 服从分布 F ，若随机变量序列 X_1, X_2, \dots, X_n 具有同一分布 F 且相互独立，则称这一随机变量序列为从总体 F 或总体 X 得到的容量为 n 的样本， x_1, x_2, \dots, x_n 为 X 的 n 个独立观测值。

反之，若一随机变量序列是总体 F 的一个样本，则序列中的随机变量同分布为 F 且相互独立。

Definition 6.2 (经验分布函数)：有样本 x_1, x_2, \dots, x_n ，用 $S(x)$ ， $-\infty < z < \infty$ 表示 x_1, x_2, \dots, x_n 中不大于 x 的随机变量的个数，定义经验分布函数 $F(z)$ 为

$$F_{n(x)} = \frac{1}{n} S(x), \quad -\infty < x < \infty$$

6.1.1 统计量

Definition 6.3 (统计量与统计量的观测值)：若有一随机变量序列 X_1, X_2, \dots, X_n 是总体 F 的一个容量为 n 的样本，则称不含有位置参数的函数 $g(X_1, X_2, \dots, X_n)$ 为统计量。

由定义可知， $g(X_1, X_2, \dots, X_n)$ 也是一个随机变量，若有 x_1, x_2, \dots, x_n 是样本的观测值，则 $g(x_1, x_2, \dots, x_n)$ 是随机变量 $g(X_1, X_2, \dots, X_n)$ 的观测值。

有总体 $X, E(X) = \mu, D(X) = \sigma^2$ ，下方为常见的统计量：

Definition 6.4 (样本平均值) : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

根据定义可得 $E(\bar{X}) = \mu, D(\bar{X}) = \frac{\sigma^2}{n}$

Definition 6.5 (样本方差) : $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$

根据定义可得, $E(S^2) = D(X) = \sigma^2$

Definition 6.6 (样本标准差) : $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$

Definition 6.7 (样本 k 阶原点矩) : $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, 3, \dots$

Definition 6.8 (样本 k 阶中心矩) : $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k = 2, 3, \dots$

6.2 抽样分布

抽样分布即为统计量为 $g(X_1, X_2, \dots, X_n)$ 的分布, 在做题时题目一般会给出提示数据, 可以查表求解。

6.2.1 χ^2 分布, t 分布和 F 分布

Definition 6.9 (χ^2 分布) : 设样本 X_1, X_2, \dots, X_n 相互独立, 且均服从 $N(0, 1)$ 分布, 则有 $X = X_1^2 + X_2^2 + \dots + X_n^2$ 服从自由度为 n 的 χ^2 分布, 即 $X \sim \chi^2(n)$ 。

χ^2 分布有如下几条性质:

Properties 6.1:

1. 可加性

若 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$ 则
 $X + Y \sim \chi^2(n_1 + n_2)$.

2. 均值与方差

若 $X \sim \chi^2(n)$, 则 $E(X) = n, D(X) = 2n$.

3. 上 α 分位点

在 χ^2 分布的密度图形中, 当 $x = x_\alpha$ 时, $x > x_\alpha$ 的面积为 α , 称此点为上 α 分位点。此时有
 $P\{X > x_\alpha\} = \alpha$.

Definition 6.10 (t 分布) : 若有 $X \sim N(0, 1), Y \sim \chi^2(n)$ 且相互独立, 则

$$\frac{X}{\sqrt{\frac{Y}{n}}} = t \sim t(n)$$

Definition 6.11 (F 分布) : 若有 $X_1 \sim \chi^2(n_1), X_2 \sim \chi^2(n_2)$ 且相互独立, 则

$$\frac{X_1/n_1}{X_2/n_2} = F \sim F(n_1, n_2)$$

与 χ^2 分布类似的, t 分布及 F 分布都具有上 α 分位点。

Definition 6.12 (正态总体的样本均值与样本方差的分布)：设总体 $X \sim N(\mu, \sigma^2)$, X_1, X_2, \dots, X_n 为总体的一个样本, 则

1. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
3. \bar{X} 与 S^2 独立
4. $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

7 参数估计

7.1 点估计

Definition 7.1：已知总体 X 的分布, 含有未知参数 θ , 用样本做参数来构造统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 来估计 θ 。

由一阶矩估计 (点估计) 推广到 k 阶矩估计, 由大数定理可得, 当数量足够大时, 样本矩趋近于总体矩, 根据矩估计中用样本矩代替总体矩的思想, 由总体的分布可以得到总体矩, 接着用样本矩代替总体矩, 也即构造未知参数 θ 与样本矩的等价关系。最后解得 $\hat{\theta}$ 即为矩估计量。

7.1.1 最大似然估计

基本思想是使得样本发生的概率最大的 $\hat{\theta}$ 即为最大似然估计。

在最大似然估计中, 用似然函数去刻画样本出现的概率大小, 对于离散型随机变量, 其最大似然函数即为样本间质量函数的积, 基本思想是使得样本发生的概率最大的 $\hat{\theta}$ 即为最大似然估计。

在最大似然估计中, 用似然函数去刻画样本出现的概率大小, 其形式如下:

$$L(\theta) = \prod_{i=0}^n P\{X = X_i\} \quad (\text{离散型随机变量})$$

$$\prod_{i=0}^n f(x_i) \quad (\text{连续型随机变量})$$

在求解最大似然估计时, 一般通过求导求其导数的驻点来得到 $\hat{\theta}$, 对于连乘函数形式的似然函数而言, 可以先等式两边同时取 **对数** 使连乘变为连加, 再求导求驻点即 $\frac{d \ln Li(\theta)}{d\theta} \triangleq 0$ 。

7.2 评选标准

1. 无偏性, 若 $E(\hat{\theta}) = \theta$ 则称 $\hat{\theta}$ 为无偏估计, 若 $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$ 则称为渐进无偏估计。
2. 有效性, 若对于未知参数 θ 有两个估计量 $\hat{\theta}_1$ 与 $\hat{\theta}_2$, 两者当中方差较小的估计量更有效。
3. 一致性, 若 n 趋于无穷时, 估计量以概率趋紧于未知参数, 则称估计量与为质量一致, 一般的, 若估计量的均值等于未知参数及具有无偏性, 估计量的方差趋近于零, 即具有有效性, 则满足估计量与未知参数具有一致性。

7.3 区间估计

Definition 7.2 (置信区间)：对于总体的一个未知参数 θ , 存在一个 α , 使得 $P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha$, 则称 $(\hat{\theta}_1, \hat{\theta}_2)$ 为置信区间。

在求解置信区间时, 通常先构造一个确定分布的含有参数 θ 的样本统计量, 也称作枢轴量 J , 根据其分布求出 $P\{a < J < b\} = 1 - \alpha$, 的左右端点 a, b (一般是由其分布的函数图像总结而来, 用上分位点表示), 进而解出 θ 的置信区间。

在构造枢轴量时，可以根据正态总体的样本均值与样本方差的分布来进行构造。