

# 概率论与数理统计

整理重邮常考的知识点

## 独立、相关、互斥、概率的计算

- 独立

如果事件A,B相互独立, 有 $P(AB) = P(A)P(B)$ ,  $E(AB) = E(A)E(B)$

- 互斥

如果事件A,B互斥, 有 $P(A + B) = P(A) + P(B)$ , 即 $P(AB) = 0$

独立与互斥没有任何关系

- 不相关

如果事件A, B不相关, 有 $Cov(A, B) = 0$

独立可以推出不相关, 不相关不能推出独立

- 概率中常用的计算公式

- 和事件的概率

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

- 德摩根律

$$P(\overline{AB}) = 1 - P(A + B)$$

- 概率拆分

$$P(AB) = P(A(1 - P(\overline{B}))) = P(A) - P(A\overline{B})$$

## 条件概率

- 乘法公式

在A的条件下, B发生的概率:

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

满足性质:

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

- 全概率公式

将A事件（条件）划分为多个事件 $A_i$ ,那么事件B发生的概率:

$$P(B) = \sum P(A_i)P(B|A_i)$$

- 贝叶斯公式

全概率公式的逆公式，表示已知B事件发生的概率下，在A中某一个划分下发生的概率:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum P(A_i)P(B|A_i)}$$

## 连续性随机变量的分布

- 概率密度函数

表示连续性随机变量在数轴上分布的稠密程度,有

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

性质:

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

- 概率分布函数

表示连续性随机变量在数轴左端分布的概率情况，即 $P(X \leq x)$ 的概率

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

性质:

$$P(a \leq x \leq b) = F(b) - F(a)$$

$$P(x > a) = 1 - F(a)$$

- 正态分布的分布函数特性

1.  $\Phi(a) = P(x \leq a)$  ,  $P(x > a) = 1 - \Phi(a)$

2.  $\Phi(0) = 0.5$

3.  $\Phi(-a) = 1 - \Phi(a)$

- 随机变量之间的函数关系

倘若随机变量X, Y之间存在某种函数关系 $Y = g(X)$ ,给定X的分布函数 $F_x(x)$ ,求Y的概率密度函数

$$F_y(y) = P(Y \leq y) = P(g(x) \leq y) = P(x \leq h(y)) = F_x(h(y))$$

$$f_Y(y) = F'_y(y) = F'_x(h(y))h'(y)$$

## 二元随机变量

- 二元连续性随机变量的分布函数

$$F(x, y) = \iint f(x, y) dA = P(A)$$

其中A是由x,y围成的区域，对应x和y的一组规划

- 边缘分布概率密度和分布函数  
对于 $F(x, y)$ ,

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

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

$$F_x(x) = F(x, \infty) = \int_{-\infty}^x f_x(x) dx$$

$$F_y(y) = F(\infty, y) = \int_{-\infty}^y f_y(y) dy$$

- 条件分布函数  
X在Y=y条件下的概率密度:

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

- 二元随机变量的分布函数

- $Z=X+Y$

$$f_z(z) = \int f(z-y, y) dy = \int f(x, z-x) dx$$

- $Z=\max\{X, Y\}$

$$F_z(z) = F_x(z)F_y(z)$$

- $Z=\min\{X, Y\}$

$$F_z(z) = 1 - [(1 - F_x(x))(1 - F_y(y))]$$

## 统计特征

- 数学期望  
离散型：略  
连续型：

$$E(x) = \int x f(x) dx$$

$$E(g(x)) = \iint g(x)f(x,y)dA$$

性质:

- $E(X + Y) = E(X) + E(Y)$
- X,Y相互独立时:  $E(XY) = E(X)E(Y)$

• 方差

$$D(x) = E(x^2) - [E(x)]^2$$

$$D(x) = \Sigma[E(x - E^2(x))]$$

性质:

- $D(X \pm Y) = D(X) + D(Y) - Cov(X,Y)$

当X,Y独立的时候才有 $D(X \pm Y) = D(X) + D(Y)$  , 在使用公式之前一定要注意X,Y是否是独立的

- $D(Cx) = c^2 D(x)$

• 协方差和相关系数

- 协方差

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

协方差反映了两个随机变量的相关性, 如果 $Cov(X,Y) = 0$  , 则X,Y不相关。

相关系数是标准化的协方差:

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

- 性质

$$Cov(X,X) = D(X)$$

$$Cov(aX + bY, cX + dY) = acD(X) + (ad + bc)Cov(X,Y) + bdD(Y)$$

• 方差矩阵

- 原点矩

指  $E(X^k)$

- 中心矩

指  $E[(X - E(X))^k]$

## 中心极限定理

中心极限定理的使用条件都是n比较大的时候。

- 李雅普诺夫定理

对于n个符合正态分布的随机变量 $X_1, X_2, \dots, X_n$ , 他们的和也服从正态分布。

$$\frac{\sum X_i - \sum \mu_i}{\sum \sigma} \sim N(0, 1)$$

- 棣莫弗-拉普拉斯定理

当n足够大时, 二项分布可视为正态分布。

若 $X \sim B(n, p)$ ,

有 $E(X) = np, D(x) = np(1 - p)$ 。

那么可以标准化X, 有:

$$\frac{X - E(X)}{\sqrt{D(X)}} \sim N(0, 1)$$

二项分布的极限可以是正态分布, 也可以是泊松分布, 优先选择泊松分布。

## 统计样本的数值特征和分布

- 样本的数值特征

- 样本均值

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

- 样本方差

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

n-1

- 样本的分布

- $\chi^2$ 分布

如果样本 $X_1, X_2, \dots, X_n$  服从正态分布, 那么 $\chi^2 = \sum x_i^2$  服从自由度为n的 $\chi^2$ 分布。

$\chi^2$ 统计量:

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

分位点:

双侧 $\alpha$ 分位点:  $\chi_{1-\frac{\alpha}{2}}^2, \chi_{\frac{\alpha}{2}}^2$

单侧 $\alpha$ 分位点:  $\chi_{1-\alpha}^2, \chi_{\alpha}^2$

- t分布

如果样本 $X \sim N(0, 1)$ , 且 $Y \sim \chi_n^2$ , X,Y独立, 那么 $t = \frac{X}{\sqrt{\frac{Y}{n^2}}}$  服从自由度为n的t分布。

t统计量:

$$t = \frac{x - \bar{\mu}}{S}$$

分位点:

双侧 $\alpha$ 分位点:  $t_{\frac{\alpha}{2}}, t_{-\frac{\alpha}{2}}$

单侧 $\alpha$ 分位点:  $t_{\alpha}, -t_{\alpha}$

◦ F分布

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$ ,  $U, V$ 相互独立, 则称 $F = \frac{\frac{U}{N_1}}{\frac{V}{N_2}}$  服从自由度为 $(n_1, n_2)$ 的F分布。

## 样本的点估计法

• 矩估计法

设 $X$ 的概率密度函数为 $f(x: \theta_1, \dots, \theta_k)$ , 用样本1~k阶矩 ( $E(x^k)$ ) 代替总体1~k阶矩建立k个方程, 联立求解, 结果是含有 $\theta$ 的式子。

◦ 结论

$$\mu = \bar{x}, \quad \sigma^2 = B_2 = A_2 - A_1^2$$

• 最大似然估计法

设 $X$ 的概率密度函数为 $f(x: \theta_1, \dots, \theta_k)$ , 通过如下方法找出参数的估计量:

1. 求解似然函数 $L(\theta) = \prod f(x)$

2. 对似然函数取对数 $\ln(L(\theta)) = \ln(\prod f(x))$

3. 求导或者是偏导数, 使每一个结果都等于0:  $\frac{\partial \ln(L(\theta))}{\partial \theta} = \frac{\partial \ln(\prod f(x))}{\partial \theta} = 0$

4. 求解 $\theta$  特殊情况: 均匀分布估计 $a, b$ 的值, 需要设最大值最小值函数

## 区间估计

区间估计遵循以下方法:

1. 选取一个合适的统计量
2. 找到 $\alpha$ 分位点, 根据备择假设的符号来判断是单边检验还是双边检验
3. 反解出关于参数的不等式
4. 代值求出不等式, 即为置信区间

## 假设检验

假设检验遵循以下方法:

1. 提出原假设和备择假设, 备择假设中不能有等于符号
2. 找到 $\alpha$ 分位点, 根据备择假设的符号来判断是单边检验还是双边检验
3. 找到拒绝域
4. 计算统计量的值, 并与分位点的值进行比对, 看统计量是否落在了拒绝域

## 随机过程的数值特征和平稳性

• 数值特征

均值函数:  $\mu_x(t) = E(X(t))$

例如：若  $X(t) = At + B$  那么  $\mu_x(t) = E(X(t)) = tE(A) + E(B)$

自相关函数：  $R_{xx}(t_1, t_2) = E[X(t_1) \cdot X(t_2)]$

方差函数：  $D_x(t) = R_{xx}(t, t) - [\mu_x(t)]^2$

- 平稳性

验证随机过程是否具有宽平稳性需要有三步：

1. 验证该过程是否是二阶矩过程\*
2. 求均值函数是否是一个参数
3. 求自相关函数  $R(t, t + \tau)$  是否只与  $\tau$  有关

## 马尔科夫链

- 概率转移矩阵

表示状态由现态转移到n次态的概率  $P = [\text{次态} \setminus \text{现态}]$

$$P(n) = p^n$$

p表示一步转移矩阵， P(n)为n步转移矩阵

- 转移概率

$$P_{ij}(m, m+n) = P(X_{m+n} = a_j | X_m = a_i)$$

表示马尔科夫链在时刻m，状态  $a_i$  下在时刻m+n下转移到状态  $a_j$  的概率