

# 概率论与数理统计

## 一、概率公式

- 加法公式:  $P(A + B) = P(A \cup B) = P(A) + P(B) - P(AB)$
- 减法公式:  $P(A - B) = P(A\overline{B}) = P(A) - P(AB)$
- 条件概率:  $P(A|B) = \frac{P(AB)}{P(B)}, P(B) > 0$
- 相容事件:  $P(AB) > 0$
- 互斥事件:  $P(AB) = 0$
- 独立事件:  $P(AB) = P(A)P(B)$
- 分布函数:

$$- F(a) = P\{X \leq a\}$$

$$- P\{X < a\} = \lim_{x \rightarrow a^-} F(x)$$

$$- f_X(x) = \frac{dF_X(x)}{dx}$$

- 卷积函数( $Z = X + Y$ ):  $f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$
- 卷积函数( $Z = X - Y$ ):  $f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(x-z)dx$

## 二、数字特征

- 数学期望:
  - 连续型: 若  $X$  的概率密度为  $f(x)$ , 则  $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$
  - 随机变量函数:  $E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$
- 二维情形边缘期望: 若  $(X, Y)$  的联合概率密度为  $f(x, y)$ , 其边缘概率密度为:

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

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

- 方差:
  - 定义式:  $D(X) = E[(X - E(X))^2]$
  - 计算式:  $D(X) = E(X^2) - [E(X)]^2$
  - 性质:  $D(CX) = C^2D(X)$
- 相关系数:  $\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}}$

- 不相关的等价命题:

$$\text{Cov}(X, Y) = 0 \iff \rho_{XY} = 0 \iff E(XY) = E(X)E(Y) \iff D(X + Y) = D(X) + D(Y)$$

- 协方差:

– 定义式:  $\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$

– 计算式:  $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

- 协方差性质

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

2.  $\text{Cov}(X, C) = 0$

3.  $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$

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

5.  $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

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

7.  $X, Y$  相互独立  $\Rightarrow \text{Cov}(X, Y) = 0$

### 三、概率分布

- 泊松分布:  $X \sim P(\lambda)$

– 结论: 若  $X \sim P(\lambda_1)$ ,  $Y \sim P(\lambda_2)$ , 且  $X$  与  $Y$  相互独立, 则  $X + Y \sim P(\lambda_1 + \lambda_2)$

- 正态分布:  $X \sim N(\mu, \sigma^2)$

– 标准化与概率:

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

– 结论: 若  $Z \sim N(0, 1)$ , 则

$$\forall a > 0, \quad P\{|Z| \leq a\} = 2\Phi(a) - 1$$

- 指数分布:  $X \sim E(\lambda)$  ( $\lambda > 0$ )

– 结论:

\*  $P\{X > a\} = e^{-\lambda a} \quad (a > 0)$

\*  $P\{X > s + t \mid X > s\} = P\{X > t\}$ , 其中  $s, t > 0$

• 常用分布表:

分布类型	分布记号	分布律/概率密度	期望	方差
0-1分布	$X \sim b(1, p)$	$P\{X = k\} = p^k(1-p)^{1-k} \quad k = 0, 1$	$p$	$p(1-p)$
二项分布	$X \sim B(n, p)$	$P\{X = k\} = C_n^k p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$	$np$	$np(1-p)$
泊松分布	$X \sim \pi(\lambda)$	$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$	$\lambda$	$\lambda$
均匀分布	$X \sim U(a, b)$	$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
正态分布	$X \sim N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < +\infty$	$\mu$	$\sigma^2$
指数分布	$X \sim E(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

## 四、数理统计

• 常用统计量:

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

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

• 三大抽样分布:

–  $\chi^2$  分布:

\* 定义:  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ , 则  $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$

\* 性质:  $E(\chi^2) = n, D(\chi^2) = 2n$

–  $t$  分布:

\* 定义:  $X \sim N(0, 1), Y \sim \chi^2(n)$  独立, 则  $T = \frac{X}{\sqrt{Y/n}} \sim t(n)$

\* 性质: 概率密度  $f(t)$  为偶函数; 若  $T \sim t(n)$ , 则  $T^2 \sim F(1, n)$

–  $F$  分布:

\* 定义:  $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$  独立, 则  $F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

\* 性质: 若  $F \sim F(n_1, n_2)$ , 则  $\frac{1}{F} \sim F(n_2, n_1)$

• 统计量的性质: 设  $X_1, X_2, \dots, X_n$  为来自总体  $X$  的样本, 且  $E(X) = \mu, D(X) = \sigma^2$ , 则:

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

- 正态总体均值、方差的检验法（显著性水平为 $\alpha$ ）：

原假设 $H_0$	检验统计量	拒绝域
$\mu \leq \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$z \geq z_\alpha$
$\mu = \mu_0$		$ z  \geq z_{\alpha/2}$
$\mu \geq \mu_0$		$z \leq -z_\alpha$
$\mu \leq \mu_0$	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$t \geq t_\alpha(n-1)$
$\mu = \mu_0$		$ t  \geq t_{\alpha/2}(n-1)$
$\mu \geq \mu_0$		$t \leq -t_\alpha(n-1)$
$\sigma^2 \leq \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2 \geq \chi_\alpha^2(n-1)$
$\sigma^2 = \sigma_0^2$		$\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)$ 或 $\chi^2 \geq \chi_{\alpha/2}^2(n-1)$
$\sigma^2 \geq \sigma_0^2$		$\chi^2 \leq \chi_{1-\alpha}^2(n-1)$

## 五、参数估计与假设检验

- 无偏性：若  $E(\hat{\theta}) = \theta$ ，则称  $\hat{\theta}$  是  $\theta$  的无偏估计量。
- 有效性：若  $\hat{\theta}_1$  和  $\hat{\theta}_2$  均为  $\theta$  的无偏估计量， $D(\hat{\theta}_1) < D(\hat{\theta}_2)$ ，则称  $\hat{\theta}_1$  比  $\hat{\theta}_2$  更有效。
- 矩估计法： $E(X) = \bar{X}$ 
  - 设  $X_1, X_2, \dots, X_n$  是来自总体的一组样本。
  - 总体矩等于样本矩： $\int_{-\infty}^{+\infty} xf(x)dx = \frac{1}{n} \sum_{i=1}^n X_i$
- 最大似然估计步骤：

1. 构造似然函数  $L(\theta) = \prod_{i=1}^n p(x_i; \theta)$

2. 取对数  $\ln L(\theta) = \sum_{i=1}^n \ln p(x_i; \theta)$

3. 解方程  $\frac{d[\ln L(\theta)]}{d\theta} = 0$

- 单正态总体  $\mu$  的置信区间（置信水平  $1 - \alpha$ ）：

1.  $\sigma^2$  已知： $\left( \bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$

2.  $\sigma^2$  未知： $\left( \bar{X} - t_{\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}} \right)$

- 假设检验中的两类错误：

- 第一类错误（弃真）：当  $H_0$  为真时，拒绝了  $H_0$ 。
- 第二类错误（取伪）：当  $H_0$  为假时，接受了  $H_0$ 。