概率论与数理统计 公式大全

一 概率论的基本概念

• 加法公式

$$egin{aligned} P(igcup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \ &\sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) + \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n) \end{aligned}$$

条件概率

$$P(B|A) = rac{P(AB)}{P(A)}$$

• 全概率公式

$$P(A) = \sum_{j=1}^n P(B_j) P(A|B_j)$$

• 贝叶斯公式

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

二 随机变量及其概率分布随机变量及其概率分布

• 概率密度

0

0

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

常见分布及期望和方差0 — 1 分布

$$P(X=k)=p^k(1-p)^{n-k}$$
 $X\sim 0-1(p), E(X)=p, D(X)=p(1-p)$

。 泊松分布

$$P(X=k) = rac{\lambda^k e^{-\lambda}}{k!} \ \ (k=0,1,2,\dots)$$
 $X \sim \pi(\lambda), E(X) = \lambda, D(X) = \lambda$

。 正态分布

0

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

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

指数分布

$$f(x) = \lambda e^{-\lambda x} \ x > 0$$

$$^{\circ} \qquad X \sim E(\lambda), E(X) = rac{1}{\mu} \, , D(X) = rac{1}{\lambda^2}$$

。 二项分布

$$P(X=k) = C_n^k \cdot P^k \cdot (1-p)^{n-k}$$
 $X \sim B(n,p), E(X) = np, D(X) = np(1-p)$

。 均匀分布

0

$$f(x) = rac{1}{b-a} \ a \leq x < b$$
 $X \sim U(a,b), E(X) = rac{a+b}{2} \, , D(X) = rac{(b-a)^2}{12}$

• 随机变量函数的概率密度

。 若
$$Y=g(x)$$
, $g'(x)>0$ 或 $g'(x)<0$

$$\circ \qquad \qquad f_Y(y) = f_x(h(y)) \cdot |h'(y)| \; lpha < y < eta$$

 \circ h(y) 是 g(x) 的反函数

三 二元随机变量

• 当 X 与 Y 相互独立时,若 Z = X + Y

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

• 二元连续型随机变量的条件概率密度

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

• **Z** = **X** + **Y**的分布

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

。 正态分布

0

$$c_0 + c_1 X_1 + c_2 X_2 + \ldots + c_n X_n \sim N(c_0 + c_1 \mu_1 + \ldots + c_n \mu_n, c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 + \ldots + c_n^2 \sigma_n^2)$$

。 二项分布

$$\circ \hspace{1cm} X+Y\sim B(n_1+n_2,p)$$

。 泊松分布

0

0

0

$$X+Y\sim\pi(\lambda_1+\lambda_2)$$

• max(X,Y)和min(X,Y)的分布

$$f_{max}(z) = f_X(z) f_Y(z)$$

$$f_{min}(z) = 1 - (1 - f_X(z))(1 - f_Y(z))$$

四 期望与方差

• 数学期望

0

0

0

0

0

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

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

$$E(Y)=E[g(X)]=\sum_{k=1}^{\infty}g(x_k)p_k$$

$$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)dx$$

$$E(Z)=E[h(X,Y)]=\sum_{i=1}^{\infty}\sum_{i=1}^{\infty}h(x_i,y_j)p_{ij}$$

$$E(Z)=E[h(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}h(x,y)f(x,y)dxdy$$

$$\circ E(cX) = xE(X)$$
 $\circ E(XY) = E(X)E(Y)$ (要求独立)

• 方差

$$D(X) = \sum_{i=1}^{+\infty} [x_i - E(X)]^2 p_2$$
 $O(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx$
 $O(X) = Var(X) = E\{[X - E(X)^2]\} = E(X^2) - [E(X)]^2$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$
 $O(X + Y) = D(X) + D(Y) + 2 \cdot Cov$

• 协方差

$$egin{aligned} & Cov(X,Y) = E\{[X-E(X)][Y-E(Y)]\} \ & & Cov(X,Y) = E(XY) - E(X)E(Y) \ & & Cov(aX,bY) = ab \cdot Cov(X,Y) \ & & Cov(X_1+X_2,Y) = Cov(X_1,Y) + Cov(X_2,Y) \ & & &
ho_{XY} = rac{Cov(X,Y)}{\sqrt{D(X)D(Y)}} \ & & & &
ho_{XY} = Cov(X^*,Y^*) \end{aligned}$$

• 不相关与独立

∘ 不相关: $\rho_{XY} = 0$

。 独立:

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j),\quad f(x,y)=f_X(x)f_Y(y)$$

• 常见分布及期望和方差

$$P(X=k)=p^k(1-p)^{n-k} \ X\sim 0-1(p), E(X)=p, D(X)=p(1-p)$$

。 泊松分布

$$P(X=k) = rac{\lambda^k e^{-\lambda}}{k!} \ \ (k=0,1,2,\ldots)$$
 $X \sim \pi(\lambda), E(X) = \lambda, D(X) = \lambda$

。 正态分布

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

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

。 指数分布

$$f(x) = \lambda e^{-\lambda x} \; x > 0$$

 $^{\circ} \qquad \qquad X \sim E(\lambda), E(X) = rac{1}{\mu} \ , D(X) = rac{1}{\lambda^2}$

。 二项分布

$$P(X=k) = C_n^k \cdot P^k \cdot (1-p)^{n-k} \ X \sim B(n,p), E(X) = np, D(X) = np(1-p)$$

。 均匀分布

0

0

$$f(x) = rac{1}{b-a} \ a \leq x < b$$

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

五 切比雪夫不等式,大数定律中心极限定理

• 切比雪夫不等式 Chebyshev's inequality

$$P\{|X-\mu| \geq \epsilon\} < rac{\sigma^2}{\epsilon^2}$$

• 伯努利大数定律

0

0

$$\lim_{n o +\infty} P\{|\,rac{n_A}{n}-p|\geq \epsilon\}=0$$

- 独立同分布的中心极限定理 (CLT)
 - 。 设 $X_1,X_2,\ldots,X_n,\ldots$ 相互独立且同分布, $E(X_i)=\mu,D(X_i)=\sigma^2,i=1,2,\ldots$ 则对于充分大 n 的,有

$$\sum_{i=1}^n X_i \sim^{ ext{ifl}} N(n\mu, n\sigma^2)$$

- 德莫弗-拉普拉斯定理
 - 。 即二项分布可以用正态分布逼近

$$\circ \qquad \qquad n_A \sim^{\inf \! igthigtarrow N} N(np, np(1-p))$$

六 统计量与抽样分布

• 样本均值

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

$$ar{X} \sim N(\mu, rac{\sigma^2}{n})$$

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

• 样本方差

0

$$S^2 = rac{1}{n-1} \sum_{i=1}^n (X_i - ar{X})^2$$

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

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

$$D(S^2) = rac{2\sigma^4}{n-1}$$

• 样本矩

0

0

0

$$A_k = rac{1}{n} \sum_{i=1}^n X_i^k$$

$$B_k = rac{1}{n} \sum_{i=1}^n (X_i - ar{X})^k$$

χ²分布

$$\chi^2 = \sum_{i=1}^n X_i^2$$

$$egin{array}{ll} \circ & E(\chi^2) = n \ \circ & D(\chi^2) = 2n \ \circ & Y_1 + Y_2 \sim \chi^2(n_1 + n_2) \end{array}$$

• *t*分布

$$T=rac{X}{\sqrt{(Y/n)}}\,, X\sim N(0,1), Y\sim \chi^2(n)$$

• **F** 分布

$$F = rac{X/n_1}{Y/n_2} \, , X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$$

• 两个正态总体的抽样分布

$$F = rac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = rac{S_1^2}{S_2^2} \left/ rac{\sigma_1^2}{\sigma_2^2} \sim F(n_1-1,n_2-1)
ight.$$

$$egin{split} (ar{X}-ar{Y}) - (\mu_1 - \mu_2) \ rac{\sigma_1^2}{n_1} + rac{\sigma_2^2}{n_2} \end{split} \sim N(0,1)$$

$$rac{(ar{X}-ar{Y})-\mu_1-\mu_2}{S_w\sqrt{rac{1}{n_1}+rac{1}{n_2}}} \sim t(n_1+n_2-2)$$

。其中

0

$$S_w^2 = rac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

七 参数估计

• 极大似然估计

0

$$L(heta) = \prod_{i=1} p(x_i; heta)$$

0

$$L(heta) = \prod_{i=1} f(x_i; heta)$$

$$egin{aligned} &\circ \ X \sim N(\mu,\sigma^2), \quad \hat{\mu} = ar{X}, \quad \hat{\sigma}^2 = B_2 \ &\circ \ X \sim U(a,b), \quad \hat{a} = \min\{X_1,\ldots,X_n\}, \quad \hat{b} = \max\{X_1,\ldots,X_n\} \end{aligned}$$

• 置信区间

$$\circ \ P\{ heta_L(X_1,\ldots,X_n)< heta<\hat{ heta}_U(X_1,\ldots,X_2)\}>1-lpha$$

八 假设检验

• 拟合优度检验

$$\sum_{i=1}^k rac{(n_i-np_i)^2}{np_i} \sim^{$$
المالية المالية المالية $\chi^2(k-r-1)$

 \circ n_i 为实际频数, np_i 为理论频数