概统公式总结

Siyan Dong

第 1 章　概率论的基本概念

加法公式（容斥原理）：

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j) + \sum_{1\le i<j<k\le n} P(A_iA_jA_k) + \ldots + (-1)^{n-1}P(A_1A_2\ldots A_n) \quad (1.1)$$

条件概率：

$$P(B|A) = \frac{P(AB)}{P(A)} \quad (1.2)$$

全概率公式：

$$P(A) = \sum_{j=1}^{n} P(B_j)P(A|B_j) \quad (1.3)$$

贝斯叶公式：

$$P(B_i|A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} \quad (1.4)$$

特别地，当 n=2 时：

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} \quad (1.5)$$

第 2 章　随机变量及其概率分布

概率密度与分布函数：

$$F(x) = P\{X \le x\} = \int_{-\infty}^{x} f(t)dt \quad (2.1)$$

二者是求导与积分的关系，在解决问题时注意相互转换；
常见分布及期望和方差：

0-1 分布:

$$P(X=k) = p^k(1-p)^{n-k}$$
$$X \sim 0-1(p), E(X) = p, D(X) = p(1-p) \quad (2.2)$$

泊松分布:

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}\,(k=0,1,2,\ldots)$$
$$X \sim \pi(\lambda), E(X) = \lambda, D(X) = \lambda \quad (2.3)$$

正态分布:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$X \sim N(\mu, \sigma^2), E(X) = \mu, D(X) = \sigma^2 \quad (2.4)$$

指数分布：

$$f(x) = \lambda e^{-\lambda x}, x > 0$$
$$X \sim E(\lambda), E(X) = \frac{1}{\mu}, D(X) = \frac{1}{\lambda^2} \tag{2.5}$$

二项分布：

$$P(X = k) = C_n^k \cdot p^k \cdot (1-p)^{n-k}$$
$$X \sim B(n,p), E(X) = np, D(X) = np(1-p) \tag{2.6}$$

均匀分布：

$$f(x) = \frac{1}{b-a}\ a \leq x < b$$
$$X \sim U(a,b), E(X) = \frac{a+b}{2}, D(X) = \frac{(b-a)^2}{12} \tag{2.7}$$

第 3 章　二元随机变量

分布函数：

$$F(x,y) = P\{X \leq x, Y \leq y\} = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy \tag{3.1}$$

边缘分布：

$$F_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$
$$F_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx \tag{3.2}$$

条件概率密度：

$$P\{X = i|Y = j\} = \frac{p_{ij}}{p_j} \tag{3.3}$$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \tag{3.4}$$

Z=X+Y 的分布：

$$F_Z(z) = P(Z \leq z) = \iint_{x+y\leq z} f(x,y)dxdy \tag{3.5}$$

$$f_Z(z) = \int_{-\infty}^{+\infty} f(z-y,y)dy \tag{3.6}$$

X 与 Y 独立时:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy \tag{3.7}$$

Z = XY:

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x})dx \tag{3.8}$$

$Z = \frac{X}{Y}$:

$$f_{X/Y}(z) = \int_{-\infty}^{+\infty} |x| f(x, zx)dx \tag{3.9}$$

正态分布:

$$c_0 + c_1X_1 + \ldots + c_nX_n \sim N(c_0 + c_1\mu_1 + \ldots + c_n\mu_n, c_1^2\sigma_1^2 + ...+, c_n^2\sigma_n^) \tag{3.10}$$

二项分布:

$$X + Y \sim B(n_1 + n_2, p) \tag{3.11}$$

泊松分布

$$X + Y \sim \pi(\lambda_1 + \lambda_2) \tag{3.12}$$

max(X,Y) 与 min(X,Y) 的分布：

$$\begin{aligned} F_{max}(z) &= F_X(z)F_Y(z) \\ F_{min}(z) &= 1 - (1 - F_X(z))(1 - F_Y(z)) \end{aligned} \tag{3.13}$$

第 4 章　随机变量的数字特征

4.1　数学期望

基本定义：

$$E(X) = \sum_{k=1}^{+\infty} x_k p_k \tag{4.1}$$

函数形式：

$$E(Z) = E[h(X,Y)] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} h(x_i, y_j)p_{ij} \tag{4.2}$$

积分形式：

$$E(Z) = E[h(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy \tag{4.3}$$

4.2　方差

定义式：

$$D(X) = \sum_{i=1}^{+\infty}[x_i - E(X)]^2 p_2 \tag{4.4}$$

积分形式：

$$D(X) = \int_{-\infty}^{+\infty}[x - E(X)]^2 f(x)dx \tag{4.5}$$

计算式：

$$D(X) = Var(X) = E\{[X - E(X)^2]\} = E(X^2) - [E(X)]^2 \tag{4.6}$$

常见其他性质：$D(X+Y) = D(X) + D(Y) + 2 \cdot Cov$；$E(X^*) = 0$；$D(X^*) = 1$

4.3　协方差与相关系数

协方差定义：

$$Cov(X,Y) = E\{[X - E(X)][Y - E(Y)]\tag{4.7}$$

计算：

$$Cov(X,Y) = E(XY) - E(X)E(Y) \tag{4.8}$$

相关系数：

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}} \tag{4.9}$$

第 5 章 大数定律中心极限定理

切比雪夫不等式 Chebyshev's inequality

$$P\{|X-\mu| \geq \epsilon\} < \frac{\sigma^2}{\epsilon^2} \tag{5.1}$$

伯努利大数定律

$$\lim_{n\to+\infty} P\{|\frac{n_A}{n}-p| \geq \epsilon\} = 0 \tag{5.2}$$

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

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \tag{5.3}$$

应用：$\frac{X-n\mu}{\sqrt{n}\sigma} \sim N(0,1)$

德莫弗-拉普拉斯定理—即二项分布可以用正态分布逼近:

$$n_A \sim N(np, np(1-p)) \tag{5.4}$$

第 6 章 统计量与抽样分布

6.1 样本均值

$$\begin{aligned} \bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ \bar{X} &\sim N(\mu, \frac{\sigma^2}{n}) \\ \frac{\bar{X}-\mu}{S/\sqrt{n}} &\sim t(n-1) \end{aligned} \tag{6.1}$$

6.2 样本方差

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \\ \frac{(n-1)S^2}{\sigma^2} &\sim \chi^2(n-1) \\ E(S^2) &= \sigma^2 \end{aligned} \tag{6.2}$$

6.3 t 分布

$$T = \frac{X}{\sqrt{(Y/n)}}, X \sim N(0,1), Y \sim \chi^2(n) \tag{6.3}$$

6.4 χ^2 分布

$$\begin{aligned} &\chi^2 = \sum_{i=1}^{n} X_i^2 \\ &E(\chi^2) = n \\ &D(\chi^2) = 2n \\ &Y_1 + Y_2 \sim \chi^2(n_1 + n_2) \end{aligned} \tag{6.4}$$

6.5 F 分布

$$F = \frac{X/n_1}{Y/n_2}, X \sim \chi^2(n_1), Y \sim \chi^2(n_2) \tag{6.5}$$

6.6 两个正态总体的抽样分布

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2}\Big/\frac{\sigma_1^2}{\sigma_2^2} \sim F(n_1 - 1, n_2 - 1) \tag{6.6}$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1) \tag{6.7}$$

$$\frac{(\bar{X} - \bar{Y}) - \mu_1 - \mu_2}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \tag{6.8}$$

$$\frac{(\bar{X} - \bar{Y}) - \mu_1 - \mu_2}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \tag{6.9}$$

其中：$S_w^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

第 7 章　参数估计

7.1 极大似然估计

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} p(x_i; \theta) \\ X \sim N(\mu, \sigma^2), &\quad \hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = B_2 \\ X \sim U(a, b), \quad \hat{a} =& \min\{X_1, \ldots, X_n\}, \quad \hat{b} = \max\{X_1, \ldots, X_n\} \end{aligned} \tag{7.1}$$

一般涉及到取 ln 和求极值

7.2 置信区间

$$P\{\hat{\theta}_L(X_1, \ldots, X_n) < \theta < \hat{\theta}_U(X_1, \ldots, X_2)\} \geq 1 - \alpha \tag{7.2}$$

和第六章各种分布的分位点有关