《概率论与数理统计》第十周要点

第四章随机变量的数字特征——方差与协方差

一、方差（Variance）

1. 定义与意义

- 方差是随机变量取值对其平均值（期望）的偏离程度，是衡量取值分散性的尺度。
- 定义：
$$D(X) = E\big[(X - E(X))^2\big],$$
- 标准差（Standard Deviation）为 $\sigma(X) = \sqrt{D(X)}$。
- 若 $D(X)$ 较小，表示 X 的取值集中在 $E(X)$ 附近；$D(X)$ 较大，则分散性强。

—

2. 计算公式

$$D(X) = E(X^2) - [E(X)]^2.$$

推导：

$$D(X) = E\{[X - E(X)]^2\} = E(X^2) - 2E(X)E(X) + [E(X)]^2 = E(X^2) - [E(X)]^2.$$

—

3. 离散型与连续型形式

- 离散型：
$$D(X) = \sum_{k}^{\infty} [x_k - E(X)]^2 p_k.$$
- 连续型：
$$D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)\, dx.$$

—

4. 方差的性质 设 C 为常数，X, Y 为随机变量。

1. $D(C) = 0$；
2. $D(CX) = C^2 D(X)$；
3. $D(X + C) = D(X)$；
4. $D(X + Y) = D(X) + D(Y) + 2\,E[(X - E(X))(Y - E(Y))]$；
5. 若 X, Y 独立，则 $D(X + Y) = D(X) + D(Y)$；
6. $D(X) = 0$ 的充要条件是 $P\{X = E(X)\} = 1$。

—

5. 典型例题 **例 1：标准化随机变量**若 $E(X) = \mu$，$D(X) = \sigma^2$，定义

$$X^* = \frac{X - \mu}{\sigma},$$

则 $E(X^*) = 0$，$D(X^*) = 1$。

—

例 2：二项分布的方差

若 $X \sim B(n, p)$，则可表示为 $X = X_1 + X_2 + \cdots + X_n$，其中每个 $X_i \sim B(1, p)$ 独立同分布。由 $E(X_i) = p$, $D(X_i) = p(1 - p)$，得

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

—

例 3：正态分布的方差

若 $X \sim N(\mu, \sigma^2)$，则 $E(X) = \mu$, $D(X) = \sigma^2$。当 a, b 为常数时，$Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$。

—

6. 切比雪夫不等式（Chebyshev Inequality） 若 $E(X) = \mu$, $D(X) = \sigma^2$，则对任意 $\varepsilon > 0$：

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

反面写作：

$$P\{|X - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$

该不等式给出了“偏离均值的概率”的上界估计。

当 $\varepsilon = k\sigma$ 时的形式 记 $\varepsilon = k\sigma$，其中 $k > 0$，代入切比雪夫不等式，可得：

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2},$$

等价地，

$$P\{|X-\mu| < k\sigma\} \geq 1 - \frac{1}{k^2}.$$

此结果表明：

- 至少有 $1 - \frac{1}{k^2}$ 的概率，随机变量的取值落在均值的 k 倍标准差之内；
- 数值示例：

  $k = 2 \Rightarrow P\{|X-\mu| < 2\sigma\} \geq \frac{3}{4} = 75\%$; $k = 3 \Rightarrow P\{|X-\mu| < 3\sigma\} \geq \frac{8}{9} \approx 88.9\%$.

- 若 X 是正态分布，真实概率会比上述界更高（例如正态分布的 3σ 范围约为 99.74%），因此切比雪夫界是“保守上界”。

—

二、协方差（Covariance）

1. 定义 设 (X, Y) 为二维随机变量，其数学期望存在。定义：

$$\mathrm{Cov}(X, Y) = E\big[(X - E(X))(Y - E(Y))\big].$$

协方差反映了 X, Y 之间的线性相关程度。

—

2. 计算公式

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

由期望性质得出，是计算协方差最常用的公式。

—

3. 性质

1. 对称性：$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$；
2. $\mathrm{Cov}(X, X) = D(X)$；
3. 线性性：对常数 a, b, c, d，

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y);$$

4. 若 X, Y 独立，则 $\operatorname{Cov}(X,Y)=0$；
5. 方差的推广式：

$$D(X+Y)=D(X)+D(Y)+2\operatorname{Cov}(X,Y).$$

—

4. 相关系数（Correlation Coefficient）

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

性质：

- $|\rho_{XY}|\le 1$；
- 当 $\rho_{XY}=0$ 时，称 X 与 Y 不相关（但未必独立）；
- 当 $|\rho_{XY}|=1$ 时，存在常数 a,b 使 $Y=a+bX$（完全线性相关）。

—

5. 协方差的应用与几何意义

- 若 $\operatorname{Cov}(X,Y)>0$：$X,Y$ 同向变化（正相关）；
- 若 $\operatorname{Cov}(X,Y)<0$：$X,Y$ 反向变化（负相关）；
- 若 $\operatorname{Cov}(X,Y)=0$：$X,Y$ 线性无关（但不一定独立）。

—

6. 典型例题　例 4：协方差计算设 X,Y 的联合分布为：

$$P(X,Y)=\begin{array}{c|cc} & Y=1 & Y=2 \\ \hline X=1 & 0.2 & 0.3 \\ X=2 & 0.1 & 0.4 \end{array}$$

计算 $\operatorname{Cov}(X,Y)$。

$$E(X)=1.5,\ E(Y)=1.9,\ E(XY)=2.95.$$

$$\operatorname{Cov}(X,Y)=E(XY)-E(X)E(Y)=2.95-(1.5)(1.9)=0.1.$$

说明：X,Y 正相关但关系较弱。

—

例 5：二维正态分布的协方差与相关系数若 $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，其联合密度为：

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right].$$

则

$$E(X)=\mu_1,\quad E(Y)=\mu_2,\quad \mathrm{Cov}(X,Y)=\rho\sigma_1\sigma_2.$$

特别地，若 $\rho=0$，则 X 与 Y 相互独立。

—

四、小结

- 方差：度量单个变量的分散性；
- 协方差：度量两个变量共同变化的程度；
- 相关系数：标准化后的协方差，$|\rho| \le 1$；
- 对正态分布类变量，独立 $\Leftrightarrow$ 不相关；
- 对一般分布类变量，独立 $\Rightarrow$ 不相关，但是不相关 $\not\Rightarrow$ 独立。