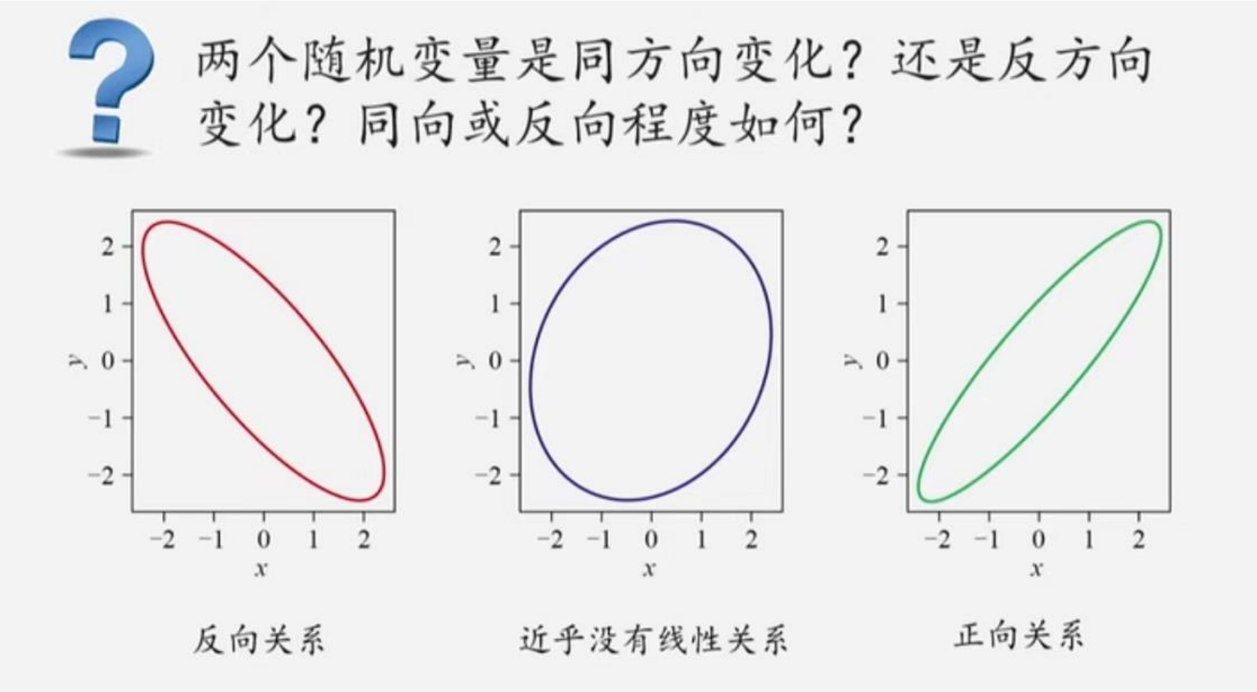
- 考虑多维随机向量的数字特征---反映**分量之间关系**的数字特征---**协方差、相关系数**

4.1. 协方差(Covariance)



定义 4.10 协方差

设随机变量 X 和 Y 均平方可积, 即 $\mathbb{E}X^2 < \infty, \mathbb{E}Y^2 < \infty$, 则称

$$\operatorname{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y), \tag{4.20}$$

为随机变量 X, Y 的协方差(covariance). ♣

$$Var(X+Y) = Var(X) + Var(Y) + 2E(X-EX)(Y-EY)$$

即

$X+Y$ **的波动性** $=X$ **的波动性** $+Y$ **的波动性** $+X$ **和** Y **的相关性**

协方差的性质

1. $Cov(X,Y) = Cov(Y,X)$, $\quad Cov(X,X) = Var(X)$。

2. $Cov(X,Y) = EXY - EXEY$。

3. 对任意实数a,b,c,d有

   $Cov(aX+b, cY+d) = acCov(X,Y)$,

   $Cov(aX+bY, cX+dY) = acVar(X) + (ad+bc)Cov(X,Y) + bdVar(Y)$.

   (随机变量作平移后的协方差不变, 随机变量前的常数可以提到协方差的外边)

4. $Cov(X_1+X_2, Y) = Cov(X_1,Y) + Cov(X_2,Y)$。

5. 若X与Y相互独立，则$Cov(X,Y) = 0$。

以概率1

6. $[\mathrm{Cov}(X,Y)]^2 \le \mathrm{Var}(X)\mathrm{Var}(Y)$. 等号成立 $\Leftrightarrow$ X, Y 之间有严格的线性关系, 即存在不全为 0 的常数 c_1, c_2, 使得 $c_1X + c_2Y + c_3 = 0$. 这条性质称为随机变量场合的柯西—施瓦茨 (Cauchy-Schwarz) 不等式.

4.2. 相关系数(correlation coefficient)

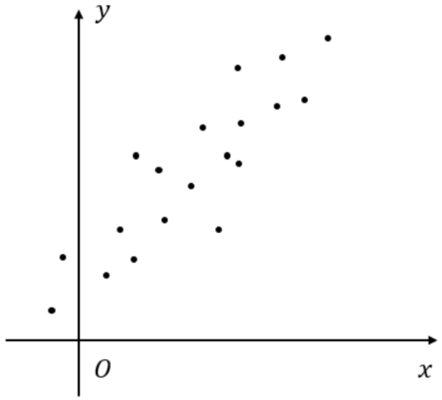
定义 4.11 相关系数

设随机变量 X 和 Y 均平方可积, 即 $\mathbb{E}X^2 < \infty, \mathbb{E}Y^2 < \infty$, 则称

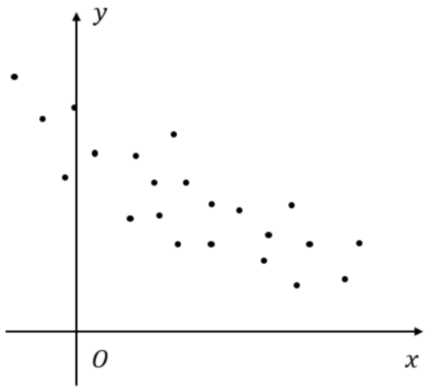
$$\rho_{X,Y} = \mathrm{Cov}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}, \frac{Y - \mathbb{E}Y}{\sqrt{\mathrm{Var}(Y)}}\right) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{4.22}$$

为随机变量 X, Y 的相关系数(correlation coefficient). 如果不混淆的话, 就简记为 ρ. ♣

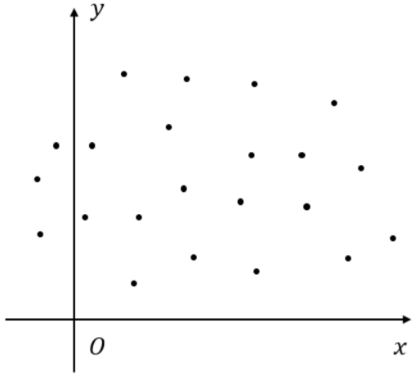
- 标准化随机变量的协方差
- 相关系数没有单位，不受随机变量单位的影响
- $|\rho_{X,Y}| \leq 1$，等号成立当且仅当X与Y之间存在严格的线性关系，即存在实数a和b使得$Y = aX + b$。
- $\rho_{X,Y} < 0$，称X, Y负相关；$\rho_{X,Y} > 0$，称X, Y正相关；$\rho_{X,Y} = 0$，称X, Y线性不相关。



(a) 正相关



(b) 负相关



(c) 线性不相关

线性相关性示意图

- **注**：$\rho_{X,Y}$也常称作 X和Y线性相关系数,只能刻画X与Y之间的**线性相依**程度。
  - ➢ $\rho_{X,Y}$越接近1，就表示X与Y之间的线性相关程度越高。
  - ➢ $\rho_{X,Y}= 0$只能表明X与Y之间**不存在线性关系**，**但可以存在非线性**的函数关系。

设 $X \sim U(-\frac{1}{2}, \frac{1}{2})$, 而 $Y = cosX$, 证明 X, Y 不相关. 但是 X, Y 之间存在着非线性的函数关系.

例子(二维正态分布)

- 设$(X,Y) \sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$，求$Cov(X,Y)$和$\rho_{X,Y}$.

独立性和不相关性

- 相互独立反映随机变量分布之间的关系，

  而不相关反映了随机变量的一个特征。

- 对于随机变量X与Y：
  - 如果X与Y相互独立，那么它们一定不相关；
  - 如果X与Y不相关，它们未必相互独立。
  - 若$(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$，则独立性与不相关性等价。

试证明若 (X, Y) 服从单位圆内的均匀分布, 则 X, Y 不相关但不独立.

设随机变量 X 和 Y 的分布律分别为

$$X \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}, \qquad Y \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

并且 $P(X \cdot Y = 0) = 1$. 则 X 与 Y 不独立, 也不相关.

定理 4.3

对任何非退化的随机变量 X, Y 存在方差, 如下四个命题相互等价: (1) X 与 Y 不相关; (2) $\mathrm{Cov}(X, Y) = 0$; (3) $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$; (4) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.