一 方差

方差衡量的是当我们对x依据它的概率分布进行采样时，随机变量x的样本值会呈现多大的差异，或者说方差是对随机变量x取值集中或分散的一种对量。

1 方差公式

$$Var(X) = E((X - E(X))^2) = E(X^2) - (E(X)^2)$$

标准差为

$\sqrt{Var(X)}$.

方差越大，随机变量的取值越分散，差异越大；方差越小，随机变量的取值越集中，差异越小。

2 方差的性质

- 设c为常数，则Var(c)=0。
- 设X为随机变量，c为常数，则:

$$Var(cX) = c^2 Var(X)\$, \$Var(c + X) = Var(X)$$

- 设X、Y是随机变量，则:

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

其实E((X−E(X)(Y−E(Y))=Cov(X,Y)，即它是随机变量X、Y的协方差。

二 协方差

协方差在某种意义上给出了两个变量线性相关的强度以及这些变量的尺度。

1 协方差公式

$$Cov(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y)$$

协方差的绝对值越大，意味着两个随机变量的线性相关性越大；协方差的绝对值越小，意味着两个随机变量的线性相关性越小。

协方差与相关性是有联系的，但只与线性相关有联系。

- 协方差为正时，两个变量是正线性相关的。
- 协方差为负时，两个变量是负线性相关的。
- 协方差为0时，两个变量不是线性相关的，但有可能是非线性相关的。

如果两个变量相互独立，那么它们的协方差一定为0，但是如果两个变量的协方差为0，这两个变量不一定相互独立，它们不是线性相关的，但有可能是非线性相关的。而相互独立不仅排除了线性相关还排除了非线性相关的可能性。

2 协方差的性质

- Cov(X,Y)=Cov(Y,X)，特别地，Cov(X,c)=0。
- Cov(aX,bY)=abCov(X,Y)。
- Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)。

三 协方差矩阵

设$\vec{X}=(X_1,X_2,\ldots,X_n)$是一个包含n个随机变量的随机向量，则$\vec{X}$的协方差矩阵是一个$n*n$的矩阵C，且：

$$C_{ij}=Cov(X_i,X_j)$$

显然矩阵C是一个对称矩阵，因为$Cov(X_i,X_j)=Cov(X_j,X_i)$，且对角线上的元素为$Cov(X_i,X_i)=Var(X_i)$。

四 相关系数(全称皮尔逊相关系数)

相关系数在协方差的基础上将每一个变量的贡献归一化，为了只衡量变量的相关性而不受各个变量尺度大小的影响。

1 相关系数公式

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

相关系数的绝对值越大，意味着两个随机变量的线性相关性越大；相关系数的绝对值越小，意味着两个随机变量的线性相关性越小。

其他性质同协方差。

2 相关系数的性质

- $\rho(X,Y)=\rho(Y,X)$。
- $|\rho(X,Y)|\leq 1$。
- $|\rho(X,Y)=1|$的充分必要条件是存在常数a,b使得$P\{Y=aX+b\}=1$,且当$a>0$时，$\rho(X,Y)=1$;当$a<0$时，$\rho(X,Y)=-1$。