概率论与数理统计——随机变量的数字特征

微信公众号：小元老师

协方差和相关系数

1 协方差

（1）定义：(X,Y)是二维随机变量，设$E(X)$和$E(Y)$都存在，若

$E\{[X-E(X)][Y-E(Y)]\}$存在，则称其为随机变量X和Y的协方差，记$Cov(X,Y)$，

即$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$.

（2）计算公式

对于任意两个随机变量X和Y，有：$Cov(X,Y)=E(XY)-E(X)E(Y)$

概率论与数理统计——随机变量的数字特征

微信公众号：小元老师

协方差和相关系数

（3）性质

1）$Cov(X,Y)=Cov(Y,X)$；

$= E(aX\cdot bY)-E(aX)\cdot E(bY)$

2）$Cov(aX,bY)=abCov(X,Y)$，其中a,b为任意常数；

3）$Cov(c,X)=0$其中c为任意常数；

4）$Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$；

5）如果X和Y相互独立，则$Cov(X,Y)=0$.

协方差和相关系数

2 相关系数

（1）定义：(X,Y)是二维随机变量，设 X 和 Y 的方差均存在，且都不为零，则称

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$ 为 X 和 Y 的（线性）相关系数.

（2）相关系数的性质

1）$|\rho_{XY}|\le 1$

2）$|\rho_{XY}|=1$ 的充分必要条件是 X 和 Y 以概率 1 线性相关，即存在常数 a 和 b，使得

$P\{Y=aX+b\}=1$，当 $a>0$ 时，$\rho_{XY}=1$；$a<0$ 时，$\rho_{XY}=-1$.

概率论与数理统计——随机变量的数字特征

微信公众号：小元老师

协方差和相关系数

$Y \backslash X$	-2	-1	1	2	$P\{Y=i\}$
1	0	1/4	1/4	0	1/2
4	1/4	0	0	1/4	1/2
$P\{X=i\}$	1/4	1/4	1/4	1/4	1

X Y.

独立
$\Downarrow \quad \not\Uparrow$
不相关

易知$E(X)=0, E(Y)=5/2, E(XY)=0$, 于是$\rho_{XY}=0$, X,Y不相关.这表示X,Y不存在线性关系.但，$P\{X=-2,Y=1\}=0\ne P\{X=-2\}P\{Y=-1\}$,知 X,Y不是相互独立的.

事实上，X 和 Y 具有关系：$Y=X^2$, Y 的值完全可由 X 的值所确定.

微信公众号：小元老师

概率论与数理统计——随机变量的数字特征

协方差和相关系数

相关不等于有因果，比如数据表明冰激凌的销售与游泳溺亡人数正相关
草莓馅饼干销量和飓风有关，概率思维不同于因果思维

相关反应线性关系，如：多吃一个汉堡，体重增加1斤

独立

$\Downarrow \quad \not\Uparrow$

不相关

微信公众号：小元老师

概率论与数理统计——随机变量的数字特征

协方差和相关系数

抛n次硬币，正面朝上和反面朝上的次数分别为X,Y，则 求X,Y的相关系数。

$X+Y=n.$

$Y=-X+n.$

$\rho_{XY}=-1.$