第四章 随机变量的数字特征

一. 协方差 *covariance*

定义 $\operatorname{cov}(X,Y)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]$

性质 1) $\operatorname{cov}(X,X)=D(X)$ $E\left[\left(X-E(X)\right)\left(X-E(X)\right)\right]$

2) $\operatorname{cov}(X,c)=E\left[\left(X-E(X)\right)\left(c-E(c)\right)\right]=0$ c为常数

3) $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$

4) $\operatorname{cov}(aX,bY)=ab\operatorname{cov}(X,Y)$ a,b 是常数

$=E\left[\left(aX-E(aX)\right)\left(bY-E(bY)\right)\right]=E\left[ab\left(X-E(X)\right)\left(Y-E(Y)\right)\right]$

5) $\operatorname{cov}(X+Y,Z)=\operatorname{cov}(X,Z)+\operatorname{cov}(Y,Z)$

6) $D(X+Y)=D(X)+D(Y)+2\operatorname{cov}(X,Y)$

方差与协方差的关系

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

证明： $D(X+Y)=E[X+Y-E(X+Y)]^2$

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

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

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

计算协方差的公式 $\operatorname{cov}(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$

$$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$$

证明： 由协方差的定义及期望的性质，

$$\operatorname{cov}(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$$

$$=E[XY-XE(Y)-YE(X)+E(X)E(Y)]$$

$$=E(XY)-E(X)E(Y)-E(Y)E(X)+E(X)E(Y)$$

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

注： 显然，若 X 与 Y 相互独立，则：$\operatorname{cov}(X, Y)=0$

例1

$X \backslash Y$	1	2	
0	0.2	0.3	0.5
1	0.4	0.1	0.5
	0.6	0.4	

求 $\operatorname{cov}(X-3, Y+2)$

$\operatorname{cov}(X+Y, X-Y)$

$D(X-Y)$

解 $\boxed{\operatorname{cov}(X-3, Y+2)}$

$$=\operatorname{cov}(X, Y)+\operatorname{cov}(X, 2)-\operatorname{cov}(3, Y)-\operatorname{cov}(3, 2)$$

$$\boxed{=\operatorname{cov}(X, Y)}$$

$$=E(XY)-E(X)\cdot E(Y)=-0.1$$

XY	0	1	2
p	0.5	0.4	0.1

$E(X)=0.5$， $E(Y)=1.4$，

$E(XY)=0.6$

$$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$$

例1

$X \backslash Y$	1	2	
0	0.2	0.3	0.5
1	0.4	0.1	0.5
	0.6	0.4	

求 $\operatorname{cov}(X-3, Y+2)$

$\operatorname{cov}(X+Y, X-Y)$

$D(X-Y)$

解　易算得：　$E(X)=0.5$，$E(Y)=1.4$，$E(XY)=0.6$

$$D(X)=E(X^2)-[E(X)]^2=0^2\cdot 0.5+1^2\cdot 0.5-0.5^2=0.25$$

同理：　$D(Y)=0.24$

$$\operatorname{cov}(X+Y, X-Y)$$

$$=\operatorname{cov}(X,X)-\operatorname{cov}(X,Y)+\operatorname{cov}(Y,X)-\operatorname{cov}(Y,Y)$$

$$=D(X)-D(Y)=0.01$$

$$D(X-Y) \quad =D(X)+D(Y)-2\operatorname{cov}(X,Y) \quad =0.69$$

二. 相关系数

$$\operatorname{cov}(X,Y)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]$$
$$=E(XY)-E(X)\cdot E(Y)$$

1. 定义2.　量(无量纲) $\dfrac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$ 称为随机变量

X, Y 的相关系数，记为：ρ_{XY}

即：$\rho_{XY}=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$

2. 相关系数的性质

(1). $|\rho_{XY}|\le 1$

(2). $|\rho_{XY}|=1 \Leftrightarrow$ 存在常数 a,b, 使得：

$$P(Y=aX+b)=1$$

X 和 Y 以概率 1 存在着线性关系

定理 设 ρ 为X与Y的相关系数

$i)$ $|\rho| \le 1$

$ii)$ $|\rho| = 1 \Leftrightarrow P\{Y = aX + b\} = 1$ a,b为常数且a不为0

若 $|\rho_{XY}| = 1$, 则 Y 与 X 以概率1存在线性关系;

若 $0 \le |\rho_{XY}| \le 1$,

$|\rho_{XY}|$的值越接近于1, Y 与 X 线性关系的程度越高;

$|\rho_{XY}|$的值越接近于0, Y 与 X 线性关系的程度越弱.

若 $\rho_{XY} = 0$, 则称 Y 与 X 不相关(没有线性关系)

相关系数刻划了X和 Y之间线性关系的紧密程度.

三. 不相关

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

定义 若 $\rho_{X,Y} = 0$，则称 X, Y 不相关

等价命题 X, Y 不相关

$\rho_{X,Y} = 0$

$\Longleftrightarrow \operatorname{cov}(X,Y) = 0$ $\not\longrightarrow$ X,Y相互独立

$\Longleftrightarrow E(XY) = E(X) \cdot E(Y)$

$$\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y)$$

$\Longleftrightarrow D(X+Y) = D(X) + D(Y)$

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

例2. 设 X,Y 在 $x^2+y^2\le 1$ 上服从均匀分布，

即：$f(x,y)=\begin{cases}\dfrac{1}{\pi} & x^2+y^2\le 1\\ 0 & x^2+y^2>1\end{cases}$

$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy$

验证：X 与 Y 是不相关的，但不是相互独立的。

$\rho_{xy}=0$

$f(x,y)\ne f_X(x)f_Y(y)$

证明：由已知，X, Y 的边缘概率密度为：

$$f_X(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2}, & |x|\le 1\\ 0, & |x|>1\end{cases}\qquad f_Y(y)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-y^2}, & |y|\le 1\\ 0, & |y|>1\end{cases}$$

显然，$f(x,y)\ne f_X(x)f_Y(y)$ 故 X 与 Y 是不独立的

$$f_X(x)=\begin{cases}\frac{2}{\pi}\sqrt{1-x^2}, & |x|\le 1\\ 0, & |x|>1\end{cases}\quad f_Y(y)=\begin{cases}\frac{2}{\pi}\sqrt{1-y^2}, & |y|\le 1\\ 0, & |y|>1\end{cases}$$

$$f(x,y)=\begin{cases}\frac{1}{\pi} & x^2+y^2\le 1\\ 0 & x^2+y^2>1\end{cases}$$

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f_X(x)dy$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\cdot f(x,y)dxdy$$

奇函数在对称区间上的积分为 0

$$E(X)=\int_{-1}^{1} x\frac{2}{\pi}\sqrt{1-x^2}\ dx=0$$

$$E(Y)=\int_{-1}^{1} y\frac{2}{\pi}\sqrt{1-y^2}\ dy=0$$

$$E(XY)=\iint\limits_{x^2+y^2\le 1} xy\frac{1}{\pi}dxdy=\frac{1}{\pi}\int_{-1}^{1}dx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} xy\ dy=0$$

从而有： $\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)=0$

于是得： $\rho_{xy}=0$　故得： X,Y 是不相关的。

X与Y不相关，但不相互独立。

独立与不相关的关系：

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

结论： X 和 Y 独立 $\longrightarrow$ $\rho_{XY} = 0$ X 与 Y 不相关

反之不一定

但对下述情形，独立与不相关等价：

若 (X, Y) 服从二维正态分布，则：$\rho_{XY} = \rho$

因此，X 与 Y 相互独立 $\Leftrightarrow$ X 与 Y 不相关

结论： 设 $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

则 X 和 Y 是相互独立 $\Longleftrightarrow$ $\rho = 0$

$$E(X)=\mu_1,\ E(Y)=\mu_2,\ D(X)=\sigma_1^2,\ D(Y)=\sigma_2^2$$

例3

$X \backslash Y$	1	2	
0	0.2	0.3	0.5
1	0.4	0.1	0.5
	0.6	0.4	

求：$\rho_{X,Y}$

XY	0	1	2
p	0.5	0.4	0.1

解　易算得：$E(X)=0.5$，$E(Y)=1.4$，

$E(XY)=0.6$

$D(X)=E(X^2)-[E(X)]^2$　$D(X)=0.25$，　$D(Y)=0.24$

$$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$$

$$\longrightarrow \operatorname{cov}(X,Y)=-0.1$$

$$\therefore \rho_{X,Y}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=-0.408$$

例4 (X, Y) 的密度为 $f(x,y)=\begin{cases} 2, & 0<x<y<1, \\ 0, & \text{其他} \end{cases}$ D

求：$\rho_{X,Y}$

解 $E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f(x,y)\,dxdy=\iint_D 2x\,dxdy=\frac{1}{3}$

$E(X^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f(x,y)\,dxdy=\iint_D 2x^2\,dxdy=\frac{1}{6}$

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

$D(X)=\frac{1}{18}$. 同理 $E(Y)=\frac{2}{3}$, $E(Y^2)=\frac{1}{2}$, $D(Y)=\frac{1}{18}$

$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy f(x,y)\,dxdy=\iint_D 2xy\,dxdy=\frac{1}{4}$

$\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)$ $\rho_{X,Y}=\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

$\mathrm{cov}(X,Y)=\frac{1}{36}$ $\Rightarrow$ $\rho_{X,Y}=\frac{1}{2}$

复习

随机变量的数字特征

	离散型随机变量	连续型随机变量
X	$E(X)=\sum_{k=1}^{\infty}x_k p_k$ ★	$E(X)=\int_{-\infty}^{+\infty}xf(x)dx$ ★
$D(X)=E[X-E(X)]^2$	$D(X)=\sum_{k=1}^{\infty}(x_k-E(X))^2 p_k$	$D(X)=\int_{-\infty}^{+\infty}(x-E(X))^2 f(x)dx$
$Y=g(X)$ g连续	$E(Y)=E[g(X)]=\sum_{k=1}^{\infty}g(x_k)p_k$	$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)dx$ ★
$Z=g(X,Y)$ g连续	$E(Z)=E[g(X,Y)]=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}g(x_i,y_j)p_{ij}$	$E(Z)=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy$ ★

复习

随机变量的数字特征

★ $E(X)$性质	$E(c)=c \quad E(cX)=cE(X) \quad E(X+Y)=E(X)+E(Y)$ X,Y独立 $\quad E(XY)=E(X)E(Y)$
★ $D(X)$性质	$D(X)=E(X^2)-[E(X)]^2 \quad D(c)=0 \quad D(cX)=c^2D(X)$ X,Y独立$,D(X+Y)=D(X)+D(Y) \quad D(X)=0 \Leftrightarrow P(X=c)=1$
★ 协方差	$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$ $=E(XY)-E(X)E(Y)$ $D(X+Y)=D(X)+D(Y)+2\operatorname{cov}(X,Y) \overset{独立}{=} D(X)+D(Y)$
相关系数	$\rho_{XY}=\dfrac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$ (1). $\lvert\rho_{XY}\rvert\le 1$ (2). $\lvert\rho_{XY}\rvert=1 \Leftrightarrow$ 存在常数 a , b,使得：$P(Y=aX+b)=1$

复习

★几种常见分布的数学期望和方差

	概率分布		$E(X)$	$D(X)$
离散型	(0−1) 分布	$X \sim B(1, p)$	p	pq
	二项分布	$X \sim B(n, p)$	np	npq
	泊松分布	$X \sim P(\lambda)$	λ	λ
连续型	均匀分布	$X \sim U(a, b)$	$(a+b)/2$	$(b-a)^2/12$
	指数分布	$X \sim Exp(\theta)$	θ	θ^2
	正态分布	$X \sim N(\mu, \sigma^2)$	μ	σ^2