§3 协方差与相关系数

讨论 r.v X、Y 之间的关系

设$(X,Y)\sim f(x,y), X\sim f_X(x), Y\sim f_Y(y)$, 则

X,Y 相互独立 $\Leftrightarrow f(x,y)=f_X(x)f_Y(y)$

问 若 X,Y不独立,如何刻画它们之间的关系?

分析 若 X,Y 相互独立, 则

$$E[(X-E(X))(Y-E(Y))]=0$$

故若

$$E[(X-E(X))(Y-E(Y))]\neq 0$$

则 X,Y必不独立, 它们之间必存在一定关系.

定义 若 X,Y的方差均存在, 记

$$\mathrm{Cov}(X,Y)\triangleq E[(X-E(X))(Y-E(Y))]$$

称$\mathrm{Cov}(X,Y)$为 X,Y 的**协方差**.

协方差的基本性质

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

1. 若 X,Y 相互独立 ➡ $\mathrm{Cov}(X,Y)=0$
2. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$
3. $D(X)=E(X-E(X))^2=\mathrm{Cov}(X,X)$
4. $$\begin{aligned}\mathrm{Cov}(X,Y)&=E[(X-E(X))(Y-E(Y))]\\&=E[XY-XE(Y)-YE(X)+E(X)E(Y)]\\&=E(XY)-E[XE(Y)]-E[YE(X)]+E(X)E(Y)\\&=E(XY)-E(X)E(Y)\end{aligned}$$

协方差的计算

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

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

5. $$\begin{aligned}D(X+Y)&=D(X)+D(Y)+2E[(X-E(X))(Y-E(Y))]\\&=D(X)+D(Y)+2\mathrm{Cov}(X,Y)\end{aligned}$$

协方差的基本性质

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

⑥ 对任意常数 a,b 有

$$\begin{aligned}\mathrm{Cov}(aX,bY)&=E[(aX-E(aX))(bY-E(bY))]\\&=abE[(X-E(X))(Y-E(Y))]\\&=ab\mathrm{Cov}(X,Y)\end{aligned}$$

⑦ $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$

定理（协方差的双线性性，bilinear property）

假设 $U=a+\sum_{i=1}^{n}b_iX_i$ 和 $V=c+\sum_{j=1}^{m}d_jY_j$，那么

$$\mathrm{Cov}(U,V)=\sum_{i=1}^{n}\sum_{j=1}^{m}b_id_j\mathrm{Cov}(X_i,Y_j).$$

协方差的意义

$\because$ X,Y相互独立 ➡ $\mathrm{Cov}(X,Y)=0$

$\therefore$ $\mathrm{Cov}(X,Y)\neq 0$ ➡ X,Y必不独立

➡ X,Y之间必存在某种“关系”

又 $\forall\, k\in R$

$$\mathrm{Cov}(kX,kY)=k^2\mathrm{Cov}(X,Y)$$

考虑“**单位化**”的r.v

$$X^*\triangleq\frac{X-E(X)}{\sqrt{D(X)}},\quad Y^*\triangleq\frac{Y-E(Y)}{\sqrt{D(Y)}}$$

定义 称

$$\rho_{XY}\triangleq\mathrm{Cov}(X^*,Y^*)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为 X,Y的**相关系数**.

问 相关系数 ρ_{XY} 刻画了 X,Y 之间什么关系？

分析 考虑 X,Y 之间的线性关系，即用r.v

$$\hat{Y}=a+bX \quad (a,b\text{为常数})$$

近似地表示r.v Y. 记**均方误差**

$$e=E[(Y-\hat{Y})^2]=E[(Y-(a+bX))^2]$$
$$=E(Y^2)+b^2E(X^2)+a^2-2bE(XY)+2abE(X)-2aE(Y)$$

令

$$\begin{cases}\dfrac{\partial e}{\partial a}=2a+2bE(X)-2E(Y)=0\\[2ex]\dfrac{\partial e}{\partial b}=2bE(X^2)-2E(XY)+2aE(X)=0\end{cases}$$

解得

$$\begin{cases}b_0=\dfrac{\mathrm{Cov}(X,Y)}{D(X)}\\[2ex]a_0=E(Y)-b_0E(X)=E(Y)-E(X)\dfrac{\mathrm{Cov}(X,Y)}{D(X)}\end{cases}$$

$$\min_{a,b} e = \min_{a,b} E[(Y-(a+bX))^2] = E[(Y-(a_0+b_0X))^2]$$

$$= E[(Y-(E(Y)-b_0E(X)+b_0X))^2]$$

$$= E\{[(Y-E(Y))-b_0(X-E(X))]^2\}$$

$$= E[(Y-E(Y))^2]+b_0^2E[(X-E(X))^2]$$

$$-2b_0E[(Y-E(Y))(X-E(X))]$$

$$= D(Y)+b_0^2D(X)-2b_0\mathrm{Cov}(X,Y)$$

$$= D(Y)+b_0\mathrm{Cov}(X,Y)-2b_0\mathrm{Cov}(X,Y)$$

$$= D(Y)-b_0\mathrm{Cov}(X,Y)$$

$$= D(Y)[1-\frac{\mathrm{Cov}^2(X,Y)}{D(X)D(Y)}]$$

$$= D(Y)(1-\rho_{XY}^2)$$

重要关系式

即有

$$\min_{a,b} E[(Y-(a+bX))^2] = E[(Y-(a_0+b_0X))^2]$$

$$= D(Y)(1-\rho_{XY}^2)$$

其中

$$\begin{cases} b_0 = \dfrac{\mathrm{Cov}(X,Y)}{D(X)} \\ a_0 = E(Y) - b_0E(X) = E(Y) - E(X)\dfrac{\mathrm{Cov}(X,Y)}{D(X)} \end{cases}$$

定理 ① $|\rho_{XY}| \le 1$

② $|\rho_{XY}| = 1 \Leftrightarrow Y \overset{a.e}{=} a + bX$（$a,b$为常数）

相关系数的实际意义

$\because \quad \min\limits_{a,b} e = \min\limits_{a,b} E[(Y-(a+bX))^2] = D(Y)(1-\rho_{XY}^2)$

$\therefore \quad |\rho_{XY}|$ 较大时均方误差 e 较小

$\Downarrow$

X,Y 之间的线性关系较密切

特别当 $|\rho_{XY}|=1$ 时，X,Y 之间几乎就是线性关系

反之，当 $|\rho_{XY}|$ 较小时，X,Y 之间的线性关系较弱

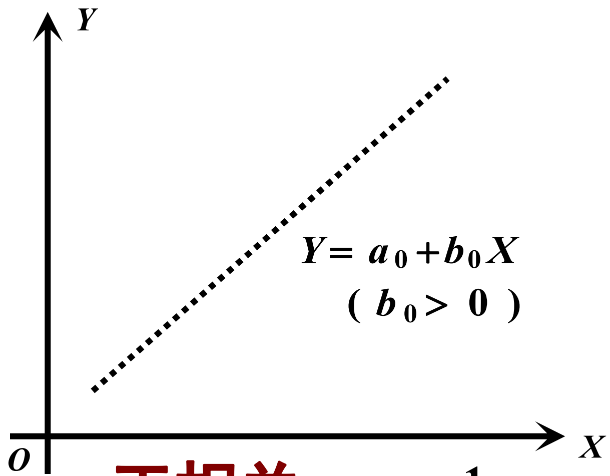
定义 当 $\rho_{XY}=1$ 时，称 X 与 Y **正相关**

当 $\rho_{XY}=-1$ 时，称 X 与 Y **负相关**

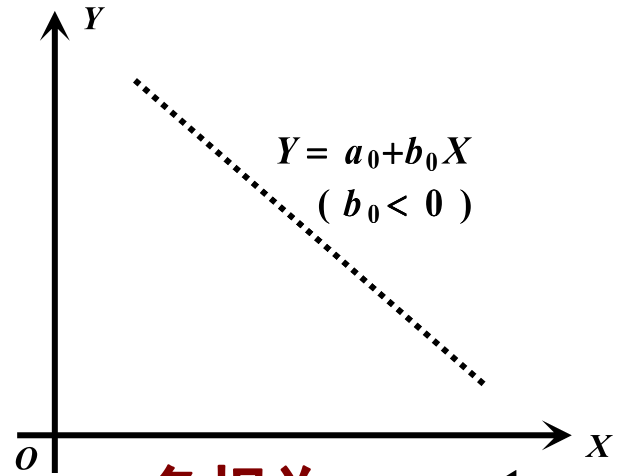
（以上两种情形统称）**相关**

当 $\rho_{XY}=0$ 时，称 X 与 Y **不相关**

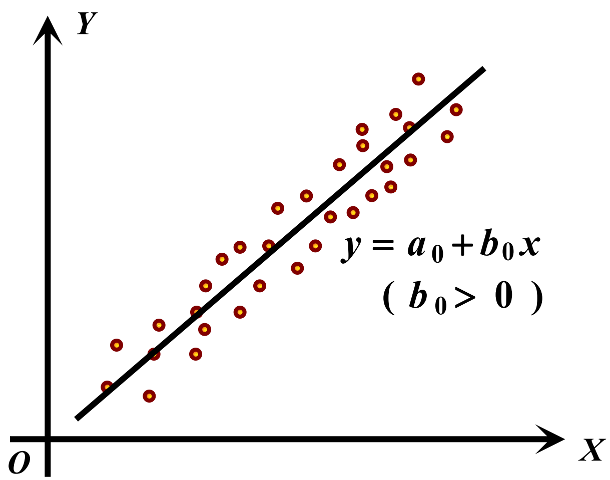
r.v X、Y 之间线性关系的图示



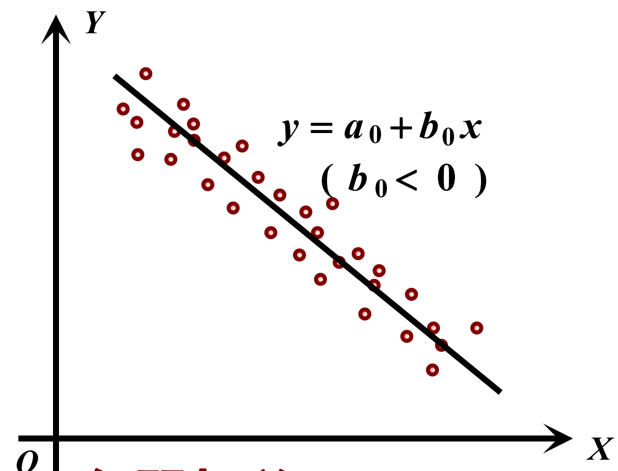
正相关 $\rho_{XY} = 1$



负相关 $\rho_{XY} = -1$

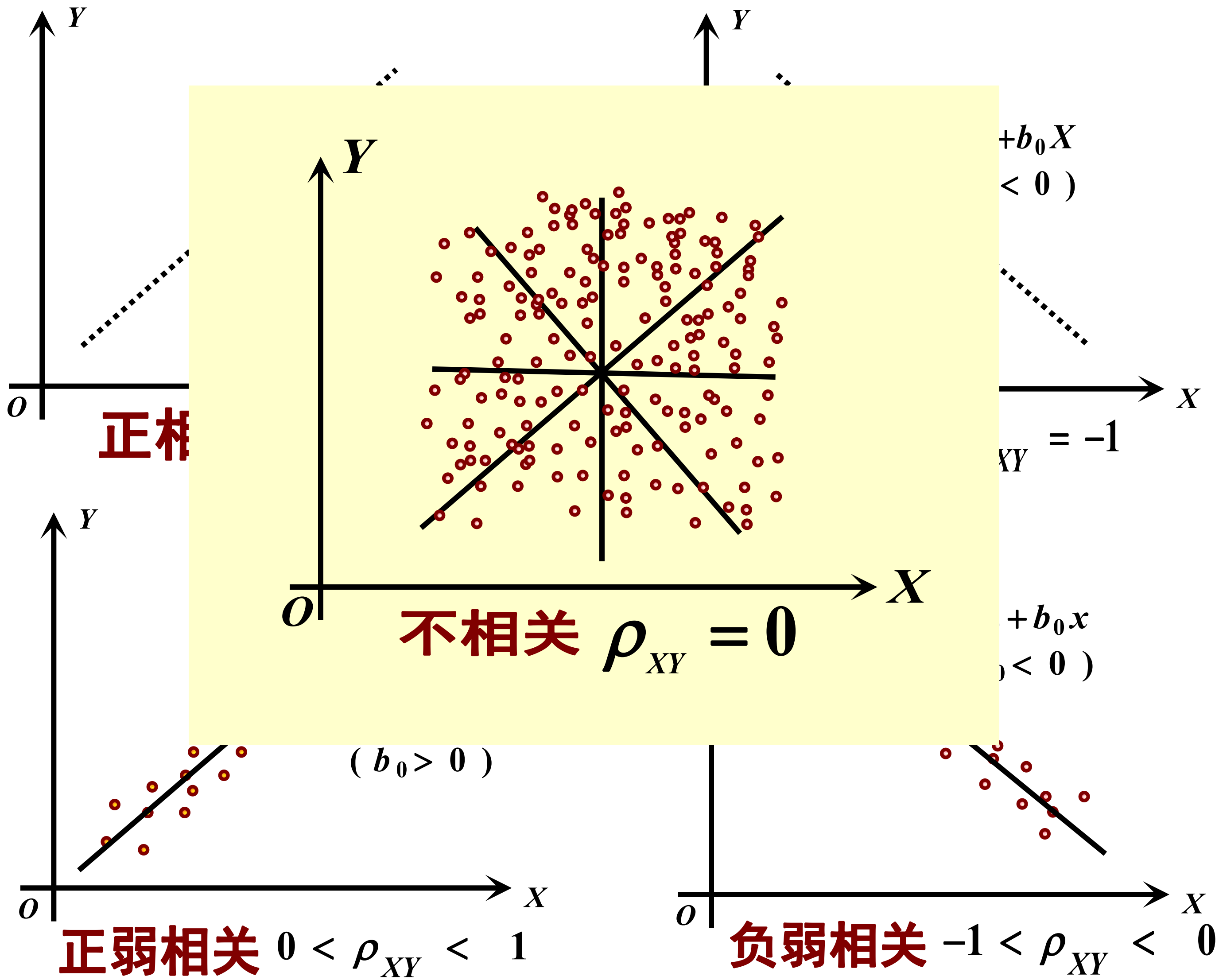


正弱相关 $0 < \rho_{XY} < 1$



负弱相关 $-1 < \rho_{XY} < 0$

r.v X、Y 之间线性关系的图示



例 设r.v (X,Y)服从圆域 $G: x^2+y^2\le 1$ 上的均匀分布，试讨论 X,Y 的独立性与相关性.

解 先求得 X,Y 的密度函数分别为

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

$$\because\ f(x,y)\ne f_X(x)f_Y(y)\quad (\forall |x|<1,\ |y|<1)$$

$\therefore$ X,Y 不独立.

又因为

$$E(X)=\int_{-\infty}^{\infty}xf_X(x)dx=0$$

$$E(Y)=\int_{-\infty}^{\infty}yf_Y(y)dy=0$$

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy=\frac{1}{\pi}\iint_{x^2+y^2\le 1}xydxdy=0$$

$$\therefore\ E[(X-E(X))(Y-E(Y))]=E(XY)-E(X)E(Y)=0$$

故 X,Y 不相关.

例 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,求 X,Y 的相关系数.

解 $\mathrm{Cov}(X,Y)=E[(X-\mu_1)(Y-\mu_2)]$

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

2. $X\sim N(\mu_1,\sigma_1^2),\ Y\sim N(\mu_2,\sigma_2^2)$

3. $E(X)=\mu_1$, $E(Y)=\mu_2$

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

猜想 $\rho=\rho_{XY}$

例 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求 X,Y 的相关系数.

解 $\mathrm{Cov}(X,Y)=E[(X-\mu_1)(Y-\mu_2)]$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-\rho^2}\,tu+\rho\sigma_1\sigma_2u^2)e^{-(t^2+u^2)/2}dtdu$$

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

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

令 $t=\frac{1}{\sqrt{1-\rho^2}}(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1})$，$u=\frac{x-\mu_1}{\sigma_1}$，$J=1$

例 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求$X,Y$的相关系数.

解 $\mathrm{Cov}(X,Y)=E[(X-\mu_1)(Y-\mu_2)]$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-\rho^2}tu+\rho\sigma_1\sigma_2u^2)e^{-(t^2+u^2)/2}dtdu$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}e^{-t^2/2}dt\cdot\int_{-\infty}^{\infty}u^2e^{-u^2/2}du$$

$$+\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}\int_{-\infty}^{\infty}te^{-t^2/2}dt\cdot\int_{-\infty}^{\infty}ue^{-u^2/2}du$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\sqrt{2\pi}\cdot\sqrt{2\pi}=\rho\sigma_1\sigma_2$$

$$\therefore\quad \rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2}=\rho$$

独立与不相关的关系

X,Y相互独立 $\substack{\Longrightarrow \\ \overset{?}{\Longleftarrow}}$ X,Y互不相关

指X,Y之间没有任何关系，当然也没有线性关系

指X,Y之间没有线性关系,但可能有其它关系

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

X,Y相互独立 $\Leftrightarrow$ $\rho=0$

$\Leftrightarrow$ $\rho_{XY}=0$

$\Leftrightarrow$ X,Y互不相关

定义 对于r.v X,Y，称

$$E(X^k) \quad (k=1,2,\cdots)$$

为k阶原点矩，简称k阶矩. 称

$$E[(X-E(X))^k] \quad (k=2,3,\cdots)$$

为k阶中心矩. 称

$$E(X^kY^l) \quad (k,l=1,2,\cdots)$$

为$k+l$阶混合矩. 称

$$E[(X-E(X))^k(Y-E(Y))^l] \quad (k,l=1,2,\cdots)$$

为$k+l$阶混合中心矩.

$E(X)$ ………… 1 阶原点矩

$D(X)$ ………… 2 阶中心矩

$\mathrm{Cov}(X,Y)$ ………… 2 阶混合中心矩

对于二维r.v (X_1, X_2)，记

$$c_{11} = E[(X_1 - E(X_1))^2] = D(X_1)$$

$$c_{12} = E[(X_1 - E(X_1))(X_2 - E(X_2))] = \mathrm{Cov}(X_1, X_2)$$

$$c_{21} = E[(X_2 - E(X_2))(X_1 - E(X_1))] = \mathrm{Cov}(X_2, X_1)$$

$$c_{22} = E[(X_2 - E(X_2))^2] = D(X_2)$$

写成矩阵的形式

$$C \triangleq \mathrm{Cov}(X, X) \triangleq \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

称矩阵 $\mathrm{Cov}(X, X)$ 为$X \triangleq (X_1, X_2)$的**协方差矩阵**.

易知 ① $C^T = C$, 即 C 为对称阵

② $C \geq 0$, 即 C 为正定(非负定)阵

证 $\Delta_1 = c_{11} = D(X_1) \geq 0$

$$\Delta_2 = |C| = c_{11}c_{22} - c_{21}c_{12}$$

$$= D(X_1)D(X_2) - [\mathrm{Cov}(X_1, X_2)]^2$$

$$= D(X_1)D(X_2)(1 - \rho_{X_1X_2}^2) \geq 0$$

课后作业

P119：54、60、补充题1，2，3

1. 设随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}\frac{1}{8}(x+y), & 0\leqslant x\leqslant 2,\ 0\leqslant y\leqslant 2,\\ 0, & \text{其他}.\end{cases}$$

求 $E(X)$, $E(Y)$, $Cov(X,Y)$, ρ_{XY}, $D(X+Y)$.

2. 设随机变量X和Y独立同分布于$N(\mu,\sigma^2)$. 令$Z=\alpha X+\beta Y$, $W=\alpha X-\beta Y$, 求$Cov(Z,W)$, ρ_{ZW}.

设随机变量X的概率密度为$f(x)=\frac{1}{2}e^{-|x|},-\infty<x<+\infty,$

(1)求出$E(X),D(X)$.

(2)X与$|X|$是否独立？说明理由.

(3)X与$|X|$是否相关？说明理由.