2. Covariance and Correlation (Rice Section 4.3, Blitzstein 7.3)

a) Define the covariance σ_{xy} and the correlation ρ_{xy} for random variables X and Y and derive the computation formula $\text{Cov}(X,Y)=E(XY)-E(X)E(Y)$. Compare to formulas for $\text{Var}(X)$.

DEFINITION

If X and Y are jointly distributed random variables with expectations μ_X and μ_Y, respectively, the covariance of X and Y is

$$\text{Cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

provided that the expectation exists. ■

DEFINITION

If X and Y are jointly distributed random variables and the variances and covariances of both X and Y exist and the variances are nonzero, then the correlation of X and Y, denoted by ρ, is

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

■

using linearity of Expectation

$\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] = E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X\mu_Y$

$= E[XY] - E[X]E[Y]$ or $E[XY] - \mu_X\mu_Y$

can be generalized.

$\text{Cov}(X,X) = E[(X-\mu_X)^2] = E[X^2] - (E[X])^2 = \text{Var}(X)$

b) As an example, consider a fat coin that has some probability of landing on its edge. Suppose the probability of it landing heads is p and the probability of it landing tails is also p $(0 < p \leq 0.5)$. If you flip the coin $n = 2$ times and let X be the number of heads and Y the number of tails, find the covariance and correlation of X and Y as a function of p, and give the values when $p = 0.4$.

$P(0<p\leq 0.5)$ flip this coin $n=2$ times, X: #heads, Y: #tails

$1-2p$

	X: H	Y: T
HH: p^2	2	0
HT: p^2	1	1
HE: $p(1-2p)$	1	0
TT: p^2	0	2
TH: p^2	1	1
TE: $p(1-2p)$	0	1
EE: $(1-2p)^2$	0	0
ET: $p(1-2p)$	0	1
EH: $p(1-2p)$	1	0

Y \ X	0	1	2	marginal:
2	p^2	$\emptyset$	$\emptyset$	p^2
1	$2p(1-2p)$	$2p^2$	$\emptyset$	$2p(1-p)$
0	$(1-2p)^2$	$2p(1-2p)$	p^2	$(1-p)^2$

X and Y are both Binomial marginally!

- $X \sim \text{Binom}(2,p)$
- $Y \sim \text{Binom}(2,p)$

$\rightarrow E[X] = np = 2p = E[Y]$

$\text{Var}[X] = npq = 2p(1-p) = \text{Var}[Y]$

$p=0.4$ $\therefore \text{Cov}(X,Y) = E[XY] - E[X]E[Y]$

$= 2p^2 - 4p^2 = -2p^2 = -0.32$.

b/c $E[XY] \neq 0$ only when $X=1$ and $Y=1$.

$\therefore \rho(X,Y) = \frac{-2p^2}{2p(1-p)} = -\frac{p}{(1-p)} = \frac{-0.4}{0.6} = -\frac{2}{3}$.

c) Derive the formulas for $\text{Var}(X+Y)$ and $\text{Var}(X-Y)$. Use the fact that $Var(X+Y)$ and $Var(X-Y)$ are both non-negative to prove that $-1 \le \rho_{xy} \le 1$ (proving either the upper or lower bound will be sufficient for the presentation).

- $\text{Var}(X+Y) = E[((X+Y)-\mu_{X+Y})^2] = E[((X-\mu_X)+(Y-\mu_Y))^2]$

$= E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] + E[2(X-\mu_X)(Y-\mu_Y)]$

$= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$

- $\text{Var}(X-Y) = E[((X-Y)-\mu_{X-Y})^2] = E[((X-\mu_X)-(Y-\mu_Y))^2]$

$= E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] - 2E[(X-\mu_X)(Y-\mu_Y)]$

$= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y)$

---

fact: $\text{Cov}(a+bX, c+dY) = bd\,\text{Cov}(X,Y)$

show $-1 \le \rho \le 1$. Suppose X, Y have mean 0.

① $\text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} + \frac{2\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$

$= 1 + 1 + 2\rho \ge 0 \rightarrow \rho \ge -1$

② $\text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} - \frac{2\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$

$= 1 + 1 - 2\rho \ge 0 \rightarrow \rho \le 1$

$\therefore -1 \le \rho \le 1.$

d) Show that independent implies uncorrelated, and explain why uncorrelated does not always imply independent. As an example, consider X and Y uniformly distributed over the triangular region $0 < |Y| < X < 1$.

Independent $\Rightarrow$ uncorrelated

$\not\Leftarrow$

($\Rightarrow$) pf)

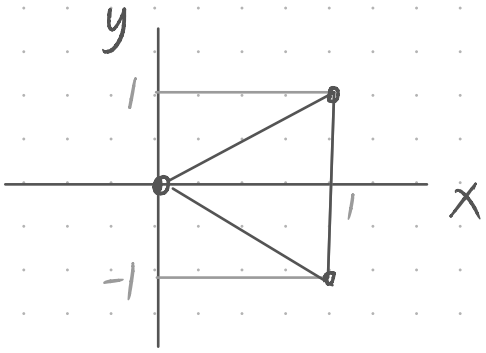
* uncorrelated iff $E[XY] = E[X]E[Y]$
* Independent iff $f_{XY}(x,y) = f_X(x)f_Y(y)$
- MV LOTUS: $E(XY) = \iint xy\, f_{XY}(x,y)\, dx\, dy$

If indep $\rightarrow$ $= \iint xy \cdot f_X(x)f_Y(y)\, dx\, dy$

$= \int x f_X(x)\, dx \cdot \int y f_Y(y)\, dy$

$= E(X)E(Y)$. $\Rightarrow$ thus, uncorrelated

(⇐) pf)



marginal pdf for X: $f_X(x) = 2x$, $0 < x < 1$

$\therefore E(X) = \int x f(x)\,dx = \frac{2}{3}$.

marginal pdf for Y is $f_Y(y) = 1 - |y|$, $-1 < y < 1$.

$\therefore E(Y) = 0$

$$\mathrm{Cov}(X,Y) = \underbrace{E(XY)}_{\downarrow} - \underbrace{E(X)E(Y)}_{\frac{2}{3}\cdot 0} = 0$$

$f(x,y) = I(0 < |y| < x < 1)$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dx\,dy = \int_{-1}^{1}\int_{|y|}^{1} xy\,dx\,dy = 0.$$

outer integral:

$$\frac{1}{2}\int_{-1}^{1} y\,dy - \frac{1}{2}\int_{-1}^{1} y^3\,dy$$

$$= \left[\frac{1}{2}y^2\right]_{-1}^{1} - \left[\frac{1}{4}y^4\right]_{-1}^{1}$$

$$= 0 - 0 = 0$$

inner integral:

$$\int_{|y|}^{1} xy\,dx = \left[\frac{x^2}{2}y\right]_{|y|}^{1}$$

$$= \frac{1}{2}y - \frac{|y|^2}{2}y$$

$$= \frac{1}{2}y - \frac{y^3}{2}$$

Thus, uncorrelated.

However, $Y|X=x \sim \mathrm{Unif}(-x, x)$ depends on X (not indep.)

$\therefore$ Uncorrelated does not always imply independent.