2.3: The correlation coefficient.

Recall :

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

$$\sigma_y = \sqrt{\frac{1}{N}\sum (y_i - \mu_y)^2} \quad , \quad \sigma_x = \sqrt{\frac{1}{N}\sum (x_i - \mu_x)^2}$$

$$\sigma_{xy} = \frac{1}{N}\sum (x_i - \mu_x)(y_i - \mu_y).$$

$$-1 \le \rho \le 1$$

Data set : $(x_1, y_1), \ldots\ldots, (x_N, y_N)$.

Def: let X_1, X_2 be Random variable with joint p.d.f $f(x_1, x_2)$. We define the covariance of X_1 and X_2 as follows :

$$\mathrm{cov}(X_1, X_2) = E\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] \text{ where } \mu_1 = E(X_1) \quad , \quad \mu_2 = E(X_2).$$

proposition : $\mathrm{cov}(X_1, X_2) = E(X_1 X_2) - \mu_1 \mu_2$.

proof :

$$\begin{aligned}
\mathrm{cov}(X_1, X_2) &= E\big((X_1 - \mu_1)(X_2 - \mu_2)\big) \\
&= E\big(X_1 X_2 - X_1 \mu_2 - X_2 \mu_1 + \mu_1 \mu_2\big) \\
&= E(X_1 X_2) - \mu_1 \mu_2 - \mu_2 \mu_1 + \mu_1 \mu_2 \\
&= E(X_1, X_2) - \mu_1 \mu_2
\end{aligned}$$

Def: let X_1, X_2 be Random variable with joint p.d.f $f(x_1, x_2)$. we define the correlation coeffecient of X_1 and X_2 as follows:

$$\rho_{xy} = \frac{cov(x_1, x_2)}{\sigma_1 \sigma_2} \quad \text{where } \sigma_1 = \sqrt{var(x_1)}, \quad \sigma_2 = \sqrt{var(x_2)}.$$

RMK: $cov(x_1, x_2)$ assume $E(x_1), E(x_2)$ exist.

ρ assume $\sqrt{var(x_1)} > 0$, $\sqrt{var(x_2)} > 0$.

Exp 1: x, y are Random variable with joint p.d.f, $\begin{cases} x+y, & 0<x<1, 0<y<1 \\ 0, & \text{elsewhere} \end{cases}$

Find ρ.

$$\rho = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{E(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} = \frac{E(xy) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

$$\rightarrow \mu_x = E(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy = \int_0^1\int_0^1 x(x+y)\,dx\,dy = \int_0^1\int_0^1 x^2+xy\,dx\,dy$$

$$= \int_0^1 \frac{x^3}{3} + \frac{x^2}{2}y \Big|_0^1 dy$$

$$= \int_0^1 \frac{1}{3} + \frac{1}{2}y\,dy$$

$$= \frac{1}{3}y + \frac{1}{2}\frac{y^2}{2}\Big|_0^1$$

$$= \frac{1}{3} + \frac{1}{4}$$

$$= \frac{7}{12}.$$

$$\rightarrow \mu_y = E(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f(x,y)\, dx\, dy = \int_0^1\int_0^1 y(x+y)\, dx\, dy = \int_0^1\int_0^1 xy + y^2\, dx\, dy$$

$$= \int_0^1 \frac{x^2}{2} y + y^2 x \Big|_0^1 \, dy$$

$$= \int_0^1 \frac{1}{2} y + y^2 \, dy$$

$$= \frac{1}{2}\frac{y^2}{2} + \frac{y^3}{3} \Big|_0^1 = \frac{7}{12}.$$

$$\rightarrow \sigma_x^2 = var(x) = E(X-\mu_x)^2 = E(x^2) - \mu_x^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\, dx\, dy - \mu_x^2$$

$$= \int_0^1\int_0^1 x^3 + x^2 y \, dx\, dy - \left(\frac{7}{12}\right)^2$$

$$= \int_0^1 \frac{x^4}{4} + \frac{x^3}{3} y \Big|_0^1 \, dy - \frac{49}{144}$$

$$= \frac{1}{4} y + \frac{1}{3}\frac{y^2}{2} \Big|_0^1 - \frac{49}{144}$$

$$= \frac{36 \times 1}{36 \times 4} + \frac{24 \times 1}{24 \times 6} - \frac{49}{144}$$

$$= \frac{36 + 24 - 49}{144}$$

$$= \frac{11}{144}$$

$\Rightarrow \delta_y^2 = Var(y) = E(y-\mu_y)^2 = E(y^2) - \mu_y^2$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f(x,y)\,dx\,dy - \mu_y^2$$

$$= \int_0^1\int_0^1 y^2x + y^3\,dx\,dy - \left(\frac{7}{12}\right)^2$$

$$\vdots$$

$$= \frac{11}{144}$$

$\Rightarrow E(xy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy = \int_0^1\int_0^1 xy\,(x+y)\,dx\,dy$

$$x^2y + xy^2$$

$$= \int_0^1 \frac{x^3}{3}y + \frac{x^2}{2}y^2 \Big|_0^1 dy$$

$$= \int_0^1 \frac{1}{3}y + \frac{1}{2}y^2\,dy$$

$$= \frac{1}{3}\frac{y^2}{2} + \frac{1}{2}\frac{y^3}{3}\Big|_0^1$$

$$= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$\Rightarrow \rho = \dfrac{E(xy) - \mu_x\mu_y}{\delta_x\delta_y} = \dfrac{\frac{1}{3} - \left(\frac{7}{12}\right)\left(\frac{7}{12}\right)}{\sqrt{\frac{11}{144}}\sqrt{\frac{11}{144}}}$

$$= \frac{-1}{11}$$

$$= -0.09$$

Remark: Assume x and y satisfy $E(y|x) = a + bx$, we can show:

$a = \mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1$ and $b = \rho \frac{\sigma_2}{\sigma_1}$

Hence, $E(y|x) = \mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1 + \rho \frac{\sigma_2}{\sigma_1} x$

$\Rightarrow E(y|x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1)$

where:

$\mu_1 = \mu_x = E(x)$.

$\mu_2 = \mu_y = E(y)$

$\sigma_1 = \sqrt{\sigma_1^2} = \sqrt{\sigma_x^2} = \sqrt{var(x)}$

$\sigma_2 = \sqrt{\sigma_2^2} = \sqrt{\sigma_y^2} = \sqrt{var(y)}$

$\rho = \frac{cov(x,y)}{\sigma_1 \sigma_2}$

similarly if $E(x|y) = c + dy$

$\Rightarrow E(x|y) = \mu_1 + \rho \frac{\sigma_1}{\sigma_2} (y - \mu_2)$

and we can show:

$E(var(y|x)) = \sigma_2^2 (1 - \rho^2)$.

Remark: $E(var(y|x)) \geq 0$

$\Rightarrow \sigma_2^2 (1 - \rho^2) \geq 0$

$\Rightarrow 1 - \rho^2 \geq 0$

$\Rightarrow -1 \leq \rho \leq 1$.

---

example 2: $E(y|x) = 4x + 3 \quad = a + bx$

$E(x|y) = \frac{1}{16} y - 3. \quad = c + dy$

1. Find μ_1, μ_2.

$\rightarrow a = 3 = \mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1$, $b = 4 = \rho \frac{\sigma_2}{\sigma_1}$

$\rightarrow c = -3 = \mu_1 - \rho \frac{\sigma_1}{\sigma_2} \mu_2$, $d = \frac{1}{16} = \rho \frac{\sigma_1}{\sigma_2}$

$\rightarrow$

→ subsitute b on a ⇒ $M_2 - 4M_1 = 3$ --- ①

→ subsitute d on c ⇒ $\left(M_1 - \frac{1}{16}M_2 = -3\right) \times 16$ --- ② } Solve

$$\cancel{M_2} - 4M_1 = 3$$
$$+ \quad 16M_1 - \cancel{M_2} = -48$$
$$12M_1 = -45 \quad \rightsquigarrow \quad \boxed{M_1 = -\frac{15}{4} = -3.75}$$

→ $M_2 = 3 + 4M_1 = 3 + 4\left(-\frac{15}{4}\right) = -12 \rightsquigarrow \boxed{M_2 = -12}$

2. Find P :

$$\underbrace{\left(P\frac{z_2}{z_1}\right)}\left(P\frac{z_1}{z_2}\right) = P^2$$

$$4 \times \frac{1}{16} = P^2$$

$$\frac{1}{4} = P^2 \rightsquigarrow \boxed{P = \frac{1}{2}}$$

$P \neq -\frac{1}{2}$ why ?

exp 3 : x

exp 4 : let X, Y have p.d.f

$$f(x,y) = \begin{cases} e^{-y} & , 0 < x < y < \infty \\ 0 & , \text{elsewhere} \end{cases}$$

calculate ρ.

$\rightarrow$ you can calculate ρ as before

$$\rho = \frac{E(xy) - E(x)E(y)}{\sqrt{E(x^2) - (E(x))^2}\sqrt{E(y^2) - (E(y))^2}}$$

- $E(x) = \int_0^{\infty}\int_0^{y} x\, f(x,y)\, dx\, dy = \int_0^{\infty}\int_0^{y} x\, e^{-y}\, dx\, dy$

- $E(y) = \int_0^{\infty}\int_0^{y} y\, e^{-y}\, dx\, dy$

- $E(x^2) = \int_0^{\infty}\int_0^{y} x^2 e^{-y}\, dx\, dy$

- $E(y^2) = \int_0^{\infty}\int_0^{y} y^2 e^{-y}\, dx\, dy$

- $E(xy) = \int_0^{\infty}\int_0^{y} xy\, e^{-y}\, dx\, dy$

Def: let X_1, X_2 be Random variables with joint p.d.f $f(X_1, X_2)$ we defined the m.g.f of X_1, X_2 are follows:

$$M(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2}).$$

if expected value exists in a Region a Round $(0,0)$.

Remark:

1. $M(0,0) = 1$ — joint m.g.f X_1, X_2 at $(t_1, t_2) = (0,0)$.

2. $M(t_1, 0) = M_1(t_1)$ — Marginal m.g.f of X_1

3. $M(0, t_2) = M_2(t_2)$ — Marginal m.g.f of X_2.

4. $\left.\dfrac{\partial^{k+L}}{\partial t_1^k \, \partial t_2^L} M(t_1, t_2)\right|_{(t_1,t_2)=(0,0)} = E(X_1^k, X_2^L) \quad k, L = 1, 2, \dots$

Back to exp 4:

lets find ρ using m.g.f

$$E(e^{t_1 x + t_2 y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f(x,y)\, dx\, dy = \int_0^{\infty}\int_0^{y} e^{t_1 x + t_2 y} e^{-y}\, dx\, dy$$

$$= \int_0^{\infty}\int_x^{\infty} e^{t_1 x + t_2 y} \quad dy\, dx$$

$$\int_0^\infty \int_0^y e^{t_1 x + t_2 y}\, e^{-y}\, dx\, dy = \int_0^\infty e^{t_2 y} e^{-y} \left(\int_0^y e^{t_1 x} dx \right) dy$$

$$= \int_0^\infty e^{(t_2-1)y} \left(\frac{e^{t_1 x}}{t_1} \Big|_0^y \right) dy$$

$$= \int_0^\infty e^{(t_2-1)y} \left(\frac{e^{t_1 y}}{t_1} - \frac{1}{t_1} \right) dy$$

$$= \int_0^\infty \frac{e^{(t_1+t_2-1)y}}{t_1} - \frac{e^{(t_2-1)y}}{t_1}\, dy$$

$$= \left(\frac{1}{t_1} \frac{e^{(t_1+t_2-1)y}}{t_1+t_2-1} - \frac{1}{t_1} \frac{e^{(t_2-1)y}}{t_2-1} \right) \Big|_0^\infty$$

$$= \left(\lim_{a\to\infty} \frac{1}{t_1} \frac{e^{(t_1+t_2-1)a}}{(t_1+t_2-1)} - \lim_{b\to\infty} \frac{1}{t_1} \frac{e^{(t_2-1)b}}{(t_2-1)} \right) - \left(\frac{1}{t_1(t_1+t_2-1)} - \frac{1}{t_1(t_2-1)} \right)$$

$$= -\frac{1}{t_1(t_1+t_2-1)} + \frac{1}{t_1(t_2-1)}$$ assuming $t_1 \neq 0$, $t_1+t_2<1$, $t_2<1$

$$= \frac{-(t_2-1) + (t_1+t_2-1)}{t_1(t_2-1)(t_1+t_2-1)}$$

$$= \frac{-\cancel{t_2} + \cancel{1} + \cancel{t_1} + \cancel{t_2} - \cancel{1}}{\cancel{t_1}(t_2-1)(t_1+t_2-1)}$$

$$= \frac{1}{(1-t_1-t_2)(1-t_2)}, \quad t_2<1,\ t_1+t_2<1$$

احلي التكامل الثاني

$dy\, dx$

$E(X) = ?$

- $M_1(t_1) = \frac{1}{1-t_1} = (1-t_1)^{-1}$, $t_1 < 1$ — لما $t=0$

- $\bar{M}_1(t_1) = (1-t_1)^{-2} \Rightarrow E(X) = 1 = \mu_1 = \mu_x$

- $\bar{\bar{M}}_1(t_1) = 2(1-t_1)^{-3} \Rightarrow E(X^2) = 2 = \sigma_x^2 = \sigma_1^2$

- $M_2(t_2) = \frac{1}{(1-t_2)^2} = (1-t_2)^{-2}$, $t_2 < 1$

- $\bar{M}_2(t_2) = 2(1-t_2)^{-3} \Rightarrow E(y) = 2 = \mu_2 = \mu_y$

- $\bar{\bar{M}}_2(t_2) = 6(1-t_2)^{-4} \Rightarrow E(y^2) = 6 = \sigma_y^2 = \sigma_2^2$

- $\left.\frac{\partial^2 M(t_1,t_2)}{\partial t_1 \, \partial t_2}\right|_{(t_1,t_2)=(0,0)} = 3 \Rightarrow E(x,y) = 3$.

$$\Rightarrow \rho = \frac{E(xy) - E(x)E(y)}{\sigma_x \sigma_y} = \frac{3-(1)(2)}{\sqrt{1}\sqrt{2}} = \frac{1}{\sqrt{2}}$$