Stat 451
2-20-18

Suppose that X & Y are independent.

①

Then $f(x,y) = g(x)\,h(y)$.

Then $E[XY] = \int\int xy\,f(x,y)\,dy\,dx$

$$= \int\int xy\,g(x)\,h(y)\,dy\,dx$$

$$= \int\left[x\,g(x)\underbrace{\int y\,h(y)\,dy}_{\mu_Y}\right]dx$$

$$= \mu_Y \int x\,g(x)\,dx = \mu_Y\,\mu_X$$

②

$\therefore$ If X & Y are independent, then

$$E[XY] = E[X]\,E[Y]$$

& so $\sigma_{XY} = 0$ and $\rho_{XY} = 0$

Note: $\rho_{XY} = 0 \not\Rightarrow$ independence

3

$$\underbrace{E\left(\left[(X-\mu_X)+v(Y-\mu_Y)\right]^2\right)}_{f(v)} \geq 0$$

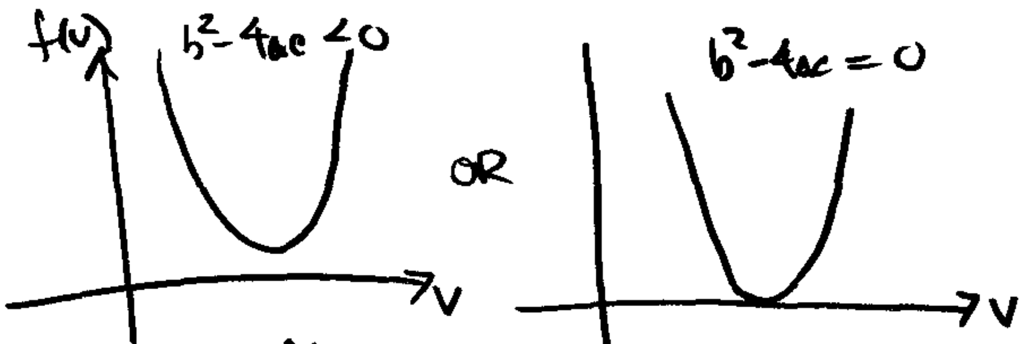
$$f(v) = E\left[(X-\mu_X)^2 + v^2(Y-\mu_Y)^2 + 2v(X-\mu_X)(Y-\mu_Y)\right]$$

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

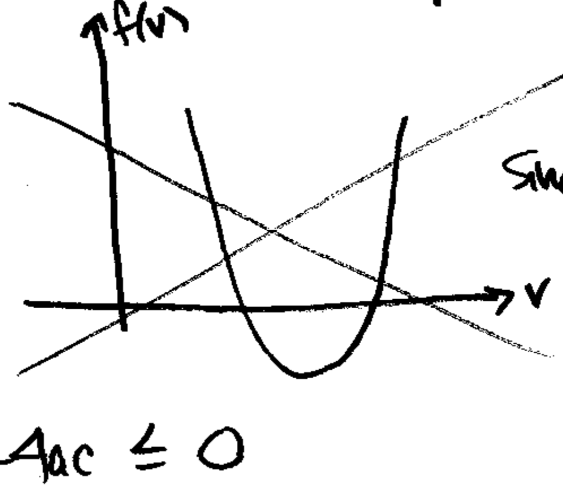
$$= \sigma_X^2 + v^2\sigma_Y^2 + 2v\,\sigma_{XY}$$

$$= \sigma_Y^2 v^2 + 2\sigma_{XY} v + \sigma_X^2$$ this is a parabola in v, opening up

4



OR



Since $f(v) \geq 0$

$\therefore b^2-4ac \leq 0$

(5)

That is, $(2\sigma_{XY})^2 - 4\sigma_Y^2\sigma_X^2 \le 0$

$$\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2 \le 0$$

$$\sigma_{XY}^2 \le \sigma_X^2\sigma_Y^2$$

$$\frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \le 1$$

$$-1 \le \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \le 1$$

$$\therefore -1 \le \rho_{XY} \le 1$$

(6)

The case where $\rho_{XY} = \pm 1$

corresponds to the case where $b^2 - 4ac = 0$

But $f(v) = E\left(\left[(X-\mu_X) + v(Y-\mu_Y)\right]^2\right)$

And it must have some v for which $f(v)=0$.

If $f(v)=0$, then

$$\left[(X-\mu_X) + v(Y-\mu_Y)\right]^2 = 0 \quad \forall X,Y$$

$$X-\mu_X + v(Y-\mu_Y) = 0 \quad \forall X,Y$$

⑦

$$X = -v(Y-\mu_Y) + \mu_X \quad \text{for some } v$$

$\therefore$ If $\rho_{XY} = \pm 1$ then X is exactly a linear function of Y

3 important ~~facts~~: ① Independence $\Rightarrow \rho_{XY} = 0$

② $|\rho_{XY}| \le 1$

③ $\rho_{XY} = \pm 1$ means perfect linear dependence between X & Y

⑧

<u>Example of correlation in a discrete distribution</u>

<u>Defn</u>: The multinomial distribution

Run independent trials.

Assume r possible outcomes on each trial, with probabilities $p_1, p_2, \dots, p_r$ such that $p_1 + p_2 + \dots + p_r = 1$.

$$\left.\begin{aligned} X_1 &= \#\text{ of type 1 outcomes} \\ &\vdots \\ X_r &= \#\text{ of type } r \text{ outcomes} \end{aligned}\right\} X_1 + \dots + X_r = n$$

9

$P(X_1 = x_1 \cap X_2 = x_2 \cap \dots \cap X_r = x_r)$

$$= \frac{n!}{x_1!\, x_2! \cdots x_r!} p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r}$$

Find the correlation between X_i and X_j

X_i is a binomial random variable with parameters n and p_i

So $E[X_i] = np_i$ and $V[X_i] = np_iq_i$
$(q_i = 1 - p_i)$

10

$X_i + X_j$ is a binomial random variable with parameters n and $(p_i + p_j)$

$$V[X_i + X_j] = n(p_i + p_j)(1 - p_i - p_j)$$

$$\parallel$$

$$V[X_i] + V[X_j] + 2\,\mathrm{Cov}(X_i, X_j)$$

$$\therefore\ \mathrm{Cov}(X_i, X_j) = \frac{n(p_i + p_j)(1 - p_i - p_j) - V[X_i] - V[X_j]}{2}$$

$$\mathrm{Cov}(X_i, X_j) = \frac{n[\cancel{p_i} + \cancel{p_j} - \cancel{p_i^2} - p_i p_j - p_i p_j - \cancel{p_j^2}] - n p_i(\cancel{1} - \cancel{p_i}) - n p_j(\cancel{1} - \cancel{p_j})}{2}$$

$$= \frac{-2n p_i p_j}{2} = -n p_i p_j$$

$$\mathrm{Corr}(X_i, X_j) = \frac{-n p_i p_j}{\sqrt{n p_i q_i}\,\sqrt{n p_j q_j}} = -\sqrt{\frac{p_i p_j}{q_i q_j}}$$