CSE 107 11-17-22

Let X, Y be ind. jointly cont. r.v.s

and

$$Z = X + Y$$

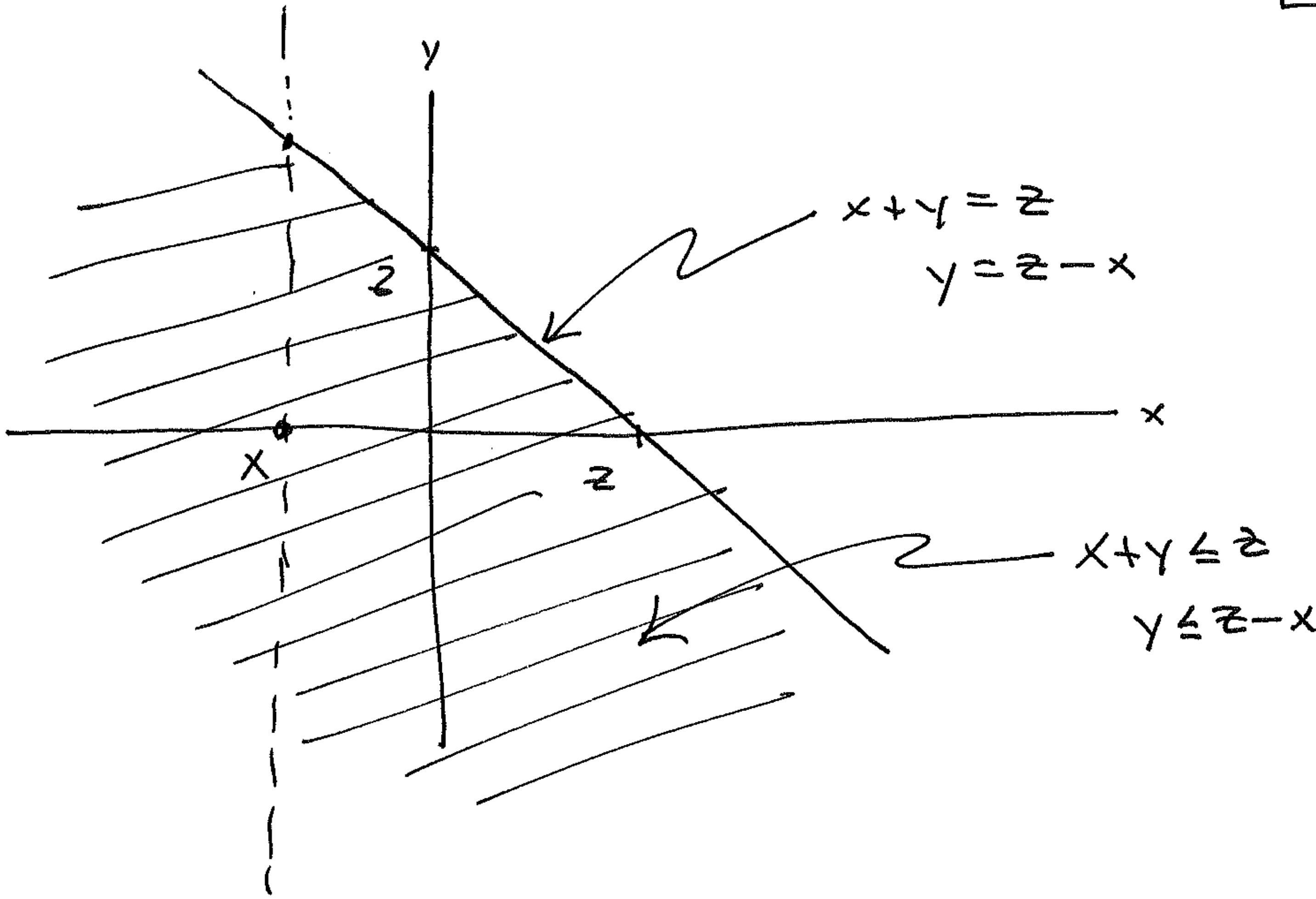
Then

$$F_Z(z) = P(Z \le z)$$

$$= P(X + Y \le z)$$

$$= \iint_{\{(x,y) \mid x+y \le z\}} f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) \cdot f_Y(y)\, dy\, dx$$



$$= \int_{-\infty}^{\infty} f_X(x) \left(\int_{-\infty}^{z-x} f_Y(y)\, dy \right) dx$$

$$= \int_{-\infty}^{\infty} f_X(x) \cdot F_Y(z-x)\, dx$$

Diff. w.r.t. z, to get

$$f_Z(z) = \frac{d}{dz}\int_{-\infty}^{\infty} f_X(x)\cdot F_Y(z-x)\,dx$$

$$= \int_{-\infty}^{\infty} f_X(x)\cdot \frac{d}{dz}F_Y(z-x)\,dx$$

$$= \int_{-\infty}^{\infty} f_X(x)\cdot f_Y(z-x)\,dx$$

This operation on f_X, f_Y is called the <u>convolution product</u>, denoted

$$f_Z = f_X * f_Y$$

$$(f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$$

Thus

$$f_{X+Y} = f_X * f_Y$$

<u>note:</u>

Both discrete & cont. versions require X, Y independent.

Ex.

let X, Y be ind. cont. uniform r.v.s on $[a,b]$. So

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Similarly for Y. Let

$$Z = X + Y.$$

Then

$$f_Z(z) = (f_X * f_Y)(z)$$

$$= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)\, dx$$

The integrand is non-zero and eq. to $\frac{1}{(b-a)^2}$ iff both

$$a \le x \le b \quad \text{and} \quad a \le z - x \le b$$

i.e.

$$a \le x \le b \quad \text{and} \quad z-b \le x \le z-a$$

i.e. iff

$$\max(a, z-b) \le x \le \min(b, z-a)$$

Thus

$$f_Z(z) = \int_{\max(a,z-b)}^{\min(b,z-a)} \frac{1}{(b-a)^2} \, dx$$

The interval of integration will be empty iff either

$$z - a < a \longrightarrow z < 2a$$

or

$$z - b > b \longrightarrow z > 2b$$

Thus we require $2a \le z \le 2b$, and

$$f_Z(z) = \begin{cases} \dfrac{\min(b, z-a) - \max(a, z-b)}{(b-a)^2} & \text{if } 2a \le z \le 2b \\ 0 & \text{otherwise} \end{cases}$$

<u>note:</u> this is not uniform.

Ex.

Let X, Y be ind. normal r.v.s with

means: μ_X, μ_Y

variances: σ_X^2, σ_Y^2

Thus

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma_X} \cdot e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}$$

similarly for Y. let

$$Z = X + Y$$

Then

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_x} \cdot e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi}\cdot\sigma_y} \cdot e^{-\frac{(z-x-\mu_y)^2}{2\sigma_y^2}}\, dx$$

(exercise)

$$= \frac{1}{\sqrt{2\pi}\cdot\sqrt{\sigma_x^2+\sigma_y^2}} \cdot e^{-\frac{(z-(\mu_x+\mu_y))^2}{2(\sigma_x^2+\sigma_y^2)}}$$

Thus Z is normal with

mean : $\mu_x + \mu_y$

Variance : $\sigma_x^2 + \sigma_y^2$

It follows that if X, Y are normal, so is

$$aX + bY.$$

4.2 Covariance & Correlation

Defn

The Covariance of r.v.s X, Y is

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

observe

$$\mathrm{Cov}(X,X) = \mathrm{var}(X)$$

- we say X, Y are <u>uncorrelated</u> iff

$$\mathrm{Cov}(X,Y) = 0$$

also

- X, Y are <u>positively correlated</u> iff

$$\mathrm{Cov}(X,Y) > 0$$

- X, Y are <u>negatively correlated</u> iff

$$\mathrm{Cov}(X,Y) < 0.$$

observe

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

$$= E\big[XY - E[X]Y - E[Y]X + E[X]E[Y]\big]$$

$$= E[XY] - E[X]E[Y] - \cancel{E[X]E[Y]} + \cancel{E[X]E[Y]}$$

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

<u>note</u>:

X, Y ind. $\Rightarrow E[XY] = E[X]E[Y]$

$\Rightarrow \mathrm{Cov}(X,Y) = 0$

$\Rightarrow X, Y$ uncorrelated.

<u>note</u>:

Converse is false. see Ex. 4.13 on P. 218–219

<u>Recall</u> contrapositive of $A \Rightarrow B$ is not $B \Rightarrow$ not A, and is logically equivalent: $(A \Rightarrow B) \equiv (\text{not } B \Rightarrow \text{not } A)$

so

$$\begin{bmatrix} X, Y \text{ Pos. or} \\ \text{neg. correlated} \end{bmatrix} \Rightarrow \begin{bmatrix} X, Y \text{ are not} \\ \text{independent} \end{bmatrix}$$

Defn

The correlation coefficient $\rho(X,Y)$

is

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}.$$

so $\rho(X,Y)$ is a normalized version of $\mathrm{Cov}(X,Y)$. We can show

$$-1 \le \rho(X,Y) \le 1$$

high neg. correlation (at -1)

high pos. correlation (at 1)

Ex.

toss a coin n times where $P(\text{head}) = p$. let

$$X = \#\text{ heads}$$

$$Y = \#\text{ tails}$$

find $\rho(X, Y)$.

Soln

observe $X + Y = n$, so

$$E[X] + E[Y] = E[X+Y] = E[n] = n.$$

and

$$Var(Y) = Var(-X+n)$$
$$= (-1)^2 Var(X) = Var(X).$$

Also

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

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

Thus

$$\begin{aligned} Cov(X,Y) &= E[-(X-E[X])^2] \\ &= -E[(X-E[X])^2] \\ &= -Var(X) \end{aligned}$$

and

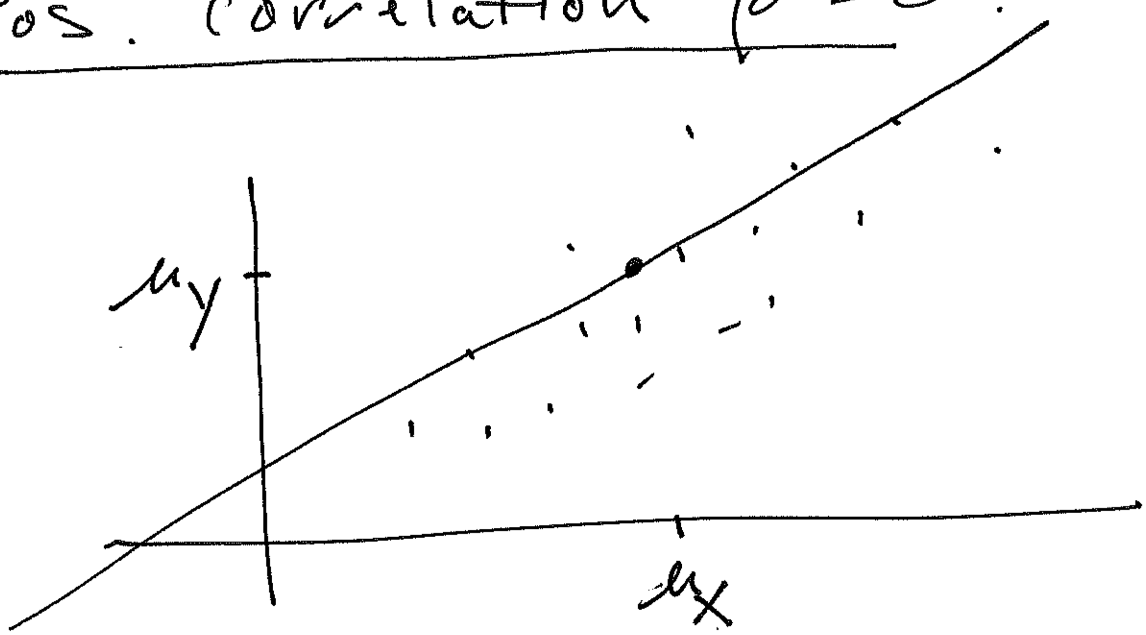
$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}}$$

$$= \frac{-\mathrm{Var}(X)}{\sqrt{\mathrm{Var}(X)^2}}$$

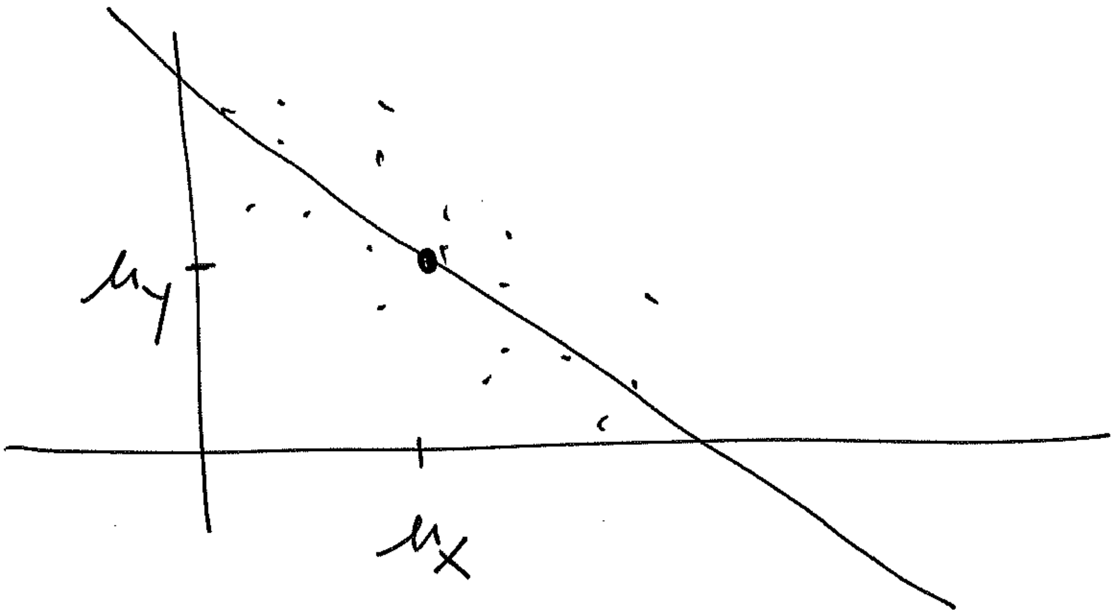
$$= -\frac{\mathrm{Var}(X)}{\mathrm{Var}(X)} = \boxed{-1}$$

as expected.

Pos. correlation $\rho > 0$:



neg. correlation $\rho < 0$:



Variance of a Sum

Let X, Y be r.v.s (not necessarily independent.) Then

$$\mathrm{Var}(X+Y)$$

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

more generally

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

Proof of case $n=2$

let

$$U = X - E[X] \therefore E[U] = 0$$

$$V = Y - E[Y] \therefore E[V] = 0.$$

$\vdots$