

Chapter 4 Expectations of functions of RVs

Example: Toss a die and win X^2
if X is # of spots.

$$\begin{aligned} E(X^2) &= \sum_{x=1}^6 x^2 p_x(x) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6} \\ &= 15 \frac{1}{6} \end{aligned}$$

Note $E(X^2) \neq (E(X))^2$

$$\begin{aligned} (E(X))^2 &= \left(1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} \right)^2 \\ &= (3.5)^2 = 12 \frac{1}{4} \end{aligned}$$

In general : $E(g(x)) = \sum_{x \in S} g(x) P_x(x)$

or

$$= \int g(x) f_x(x) dx$$

Works for function of random vectors;

Let $Y = g(X_1, X_2, \dots, X_k)$

with pmf $P(x_1, x_2, \dots, x_k)$

then $E(Y) = E(g(X_1, \dots, X_k)) = \sum_{x \in S} g(x_1, \dots, x_k) P(x_1, \dots, x_k)$

OR $E(Y) = \int \left[\int \dots \int g(x_1, \dots, x_k) f(x_1, \dots, x_k) dx_k \dots dx_2 \right] dx_1$

if the integral ^(sum) with $|g|$ converges.

P.124 Corollary A

If X & Y are independent, then

$$E(g(X)h(Y)) = E(g(X)) \times E(h(Y))$$

provided \uparrow and \uparrow exist.

Special case: If X & Y are independent

$$E(XY) = E(X)E(Y)$$

if \uparrow and \uparrow exist.

Beware: 1) not true in general
2) converse not true

Linear combinations of R.V.'s

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_n X_n$$

then $E(Y) = a + b_1 E(X_1) + b_2 E(X_2) + \dots + b_n E(X_n)$
if $\uparrow \dots \uparrow$ exist.

Example A (p. 126)

What is $E(Y)$ if $Y \sim \text{Bin}(n, p)$?

$$E(Y) = \sum_y \binom{n}{y} y p^y (1-p)^{n-y}$$

Easier: Use fact that $Y = X_1 + \dots + X_n$

where X_i 's are indep. $\text{Bin}(1, p)$

and $E(X_i) = p$

$$\text{So } E(Y) = \sum_{i=1}^n E(X_i) = np$$

Example C p. 128

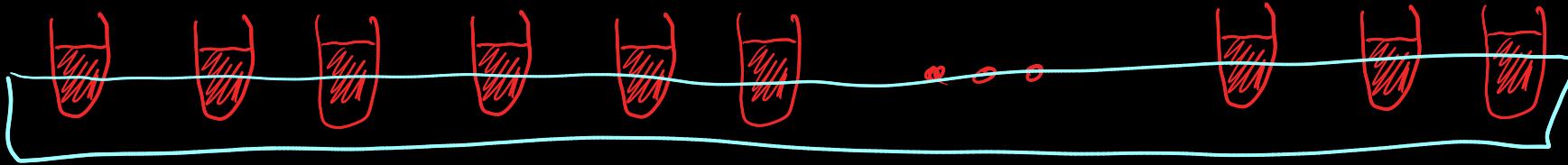
Group testing

- n blood samples tested for rare disease



- If you test each individually will need n tests.

Alternative:



take $\frac{1}{2}$ of each sample and combine:

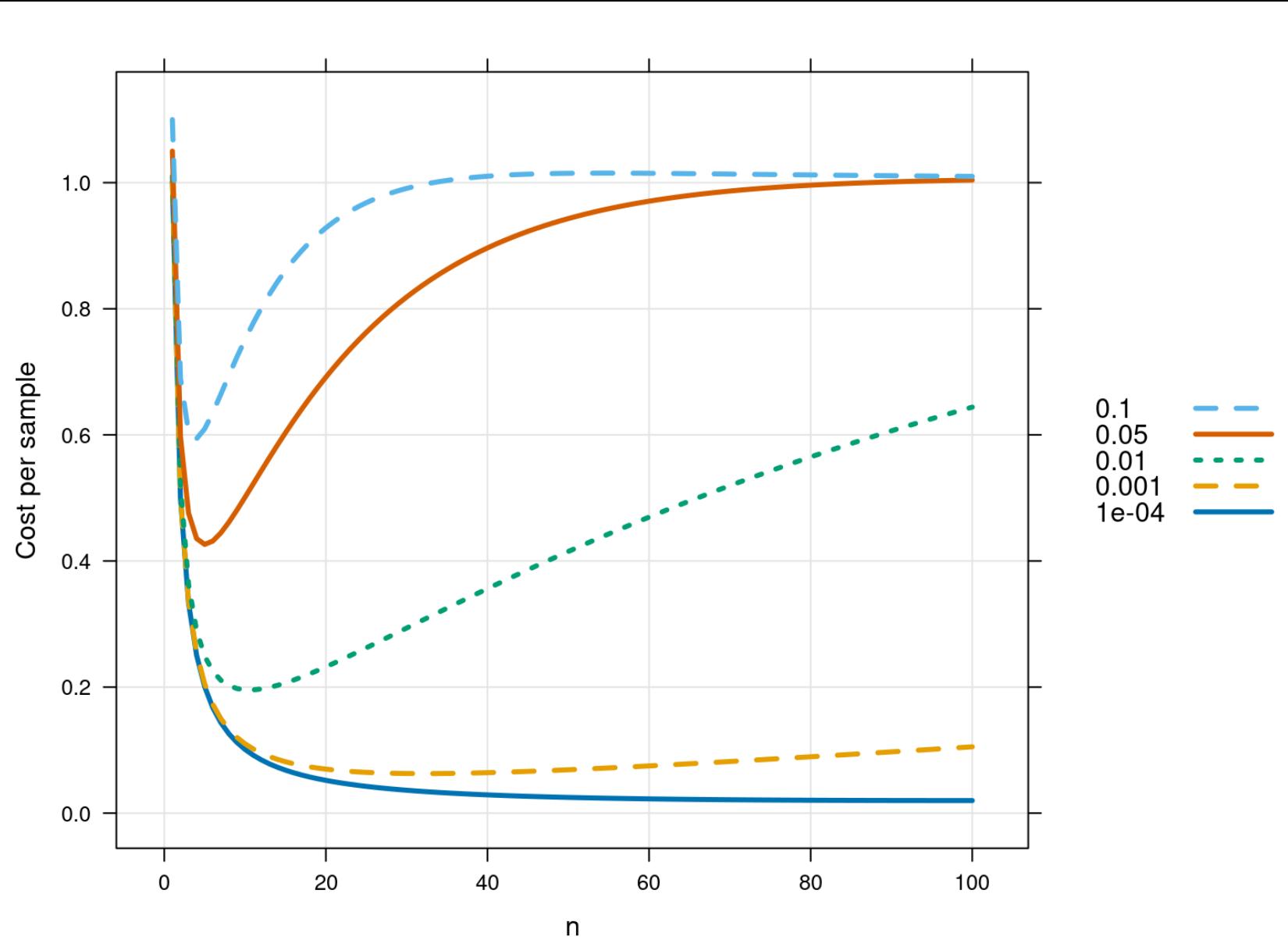


Then test.

If negative - done - all okay
If positive - test n samples.

Let p = probability of a positive.

$$E(\text{Tests}) = 1 \times \underbrace{(1-p)^n}_{\text{pr all negative}} + (n+1) \underbrace{\left(1 - (1-p)^n\right)}_{\text{prob. at least 1 positive}}$$



```
df <- expand.grid(n = 1:100, p = c(.0001,.001, .01, .05, .1))
head(df)
dim(df)
df <- within(df,
{
  Etests <- 1* (1-p)^n + (n + 1) * (1 - (1 - p)^n)
  Cost_per_sample <- Etests/n
})
library(latticeExtra)
trellis.par.set(superpose.line = list(lwd=3, lty = 1:3))
xyplot(Cost_per_sample ~ n, df,
  groups = p, type = 'l',
  ylab = "Cost per sample",
  auto.key = list(reverse.rows = T)) +
layer_(panel.grid(h=-1, v = -1))
```

Example D : Illustrates how
 $E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$
does not require independence

DNA Sequences : formed from 4 letters A C T G
of random & each letter with = prob.

ATC AATCGAGT ... TAA

- Suppose length = N
- and each letter has $P = 1/4$

How many ATGC do you expect?

Let I_n = event ATGC starts at position n

$$P(I_n) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{256}$$

$$= E(I_n)$$

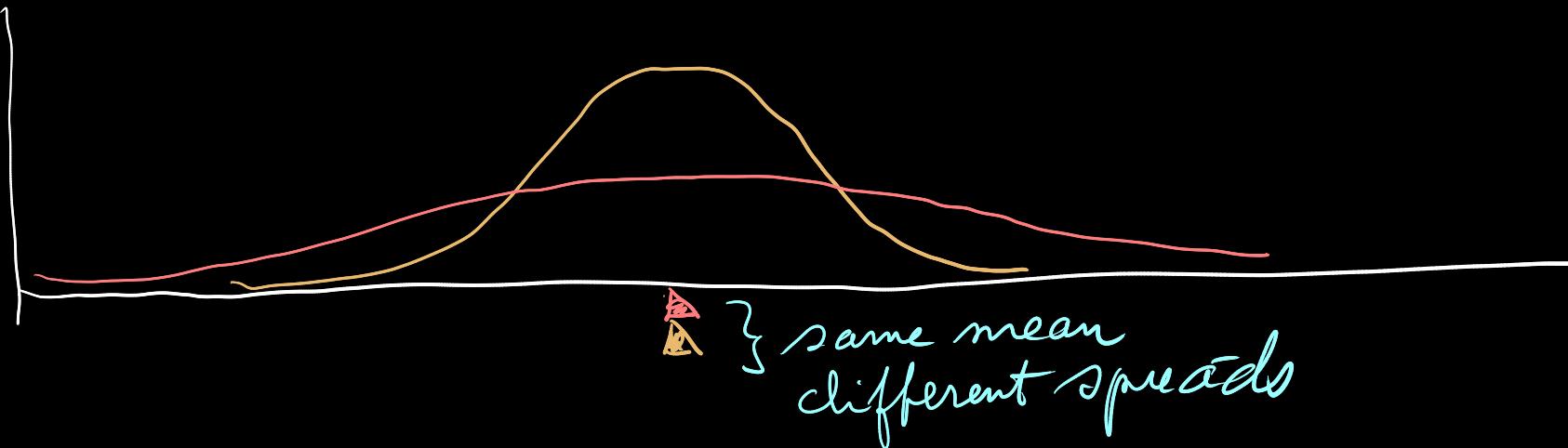
if $I_n = \begin{cases} 1 & \text{if ATGC starts} \\ 0 & \text{otherwise} \end{cases}$ at position n

$$\begin{aligned} E(\# \text{of sequences}) &= E\left(\sum_{n=1}^{N-3} I_n\right) = \sum_{n=1}^{N-3} E(I_n) \\ &= (N-3) \times \frac{1}{256} \end{aligned}$$

Variance and Standard Deviation

Mean = "location parameter" ← one of many
e.g. median, min, max

Next we need a "spread" parameter



Variance · Average squared distance from the mean

$$\text{Var}(X) = E[(X - \mu_x)^2] \quad \begin{matrix} \text{if } E \text{ exists.} \\ \text{in squared units} \end{matrix}$$

$$SD(X) = \sqrt{\text{Var}(X)} \quad \text{in original units}$$

Facts about variance.

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

Proof

$$\begin{aligned} & E[(X - \mu_x)^2] \\ &= E(X^2 - 2\mu_x X + \mu_x^2) \\ &= E(X^2) - 2\mu_x E(X) + \mu_x^2 \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

also a useful
way to
calculate
 $\text{Var}(X)$

Corollary : $E(X^2) = E(X)^2$ iff $\text{Var}(X) = 0$

Fact : $\text{Var}(X) = 0$ iff X is a constant
i.e. $P(X = c) = 1$

Fact : If $\text{Var}(X) < 0$ (i.e. exists)
and $Y = a + bX$
then $\text{Var}(Y) = b^2 \text{Var}(X)$

Chebyshov's Inequality : Prop & spread

Let X have mean μ and variance σ^2

Then for any $t > 0$:

$$P(|X - \mu| > t) \leq \sigma^2 / t^2$$

Proof: Use Markov's inequality on $Y = (X - \mu)^2$

Different form : If $\sigma > 0$, let $t = k\sigma$ for $k > 0$. Then

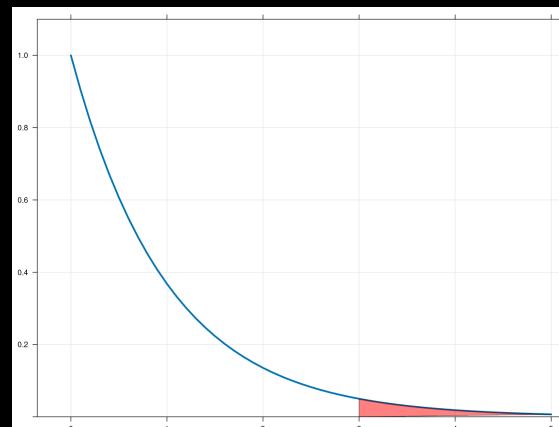
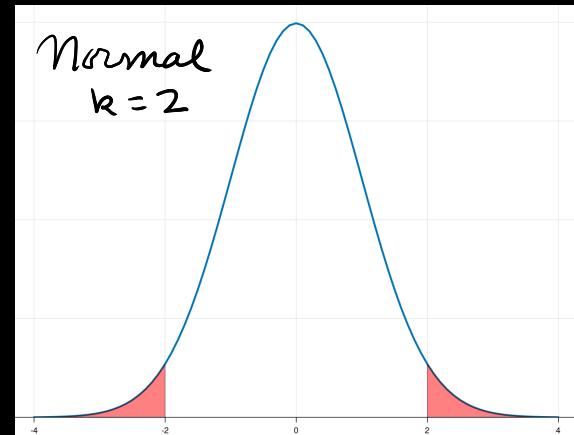
$$P(|X - \mu| > k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

For $k = 1$ this says $P(|X - \mu| \geq \sigma) \leq \frac{1}{1^2} = 1$!?

Only useful for $k > 1$; e.g. $P(|X - \mu| \geq 2\sigma) \leq \frac{1}{2^2} = \frac{1}{4}$

Let's compare with some actual values.

	$k = 2$ $\leq \frac{1}{4} = 0.25$	$k = 3$ $\leq \frac{1}{9} = 0.111\dots$	$k = 4$ $\leq \frac{1}{16} = 0.0625$
<u>Chelyshev</u>			
<u>Actuals</u>			
Normal	0.0455	0.0027	0.0000633
Exponential	0.0498	0.0183	0.00674
Poisson	0.0803	0.0196	0.00366



Fact (Corollary A, p. 134)

$$\text{Var}(X) = 0 \text{ iff } P(X = \mu) = 1$$

iff $P(X \text{ is constant}) = 1$

Proof: Example of Chebyshev's Theorem.

$$P(|X - \mu| > \varepsilon) \leq \sigma^2 / \varepsilon^2 = 0$$

for every $\varepsilon > 0$.

$$\text{So } P(|X - \mu| = 0) = 1$$

i.e. $P(X = \mu) = 1$

Investment Portfolios (p. 134)

Two investments : R_1 : $E(R_1) = \mu_1 = 0.10$
 $\sigma_1 = 0.075$

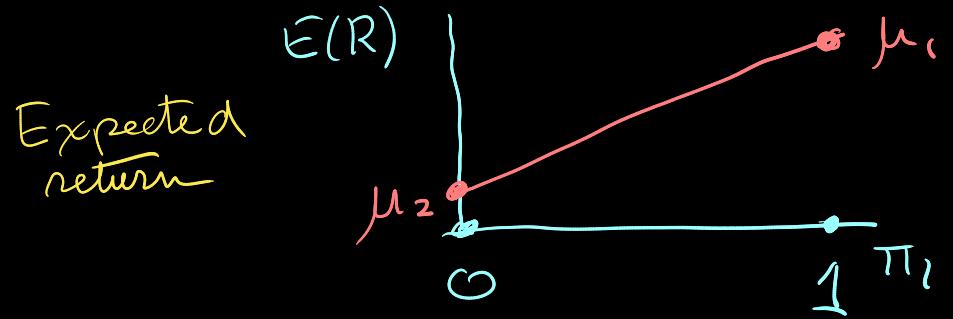
R_2 : $E(R_2) = \mu_2 = 0.03$
 $\sigma_2 = 0$

Same cost so investor can choose
to invest π_1 in R_1 and $\pi_2 = (1 - \pi_1)$ in R_2

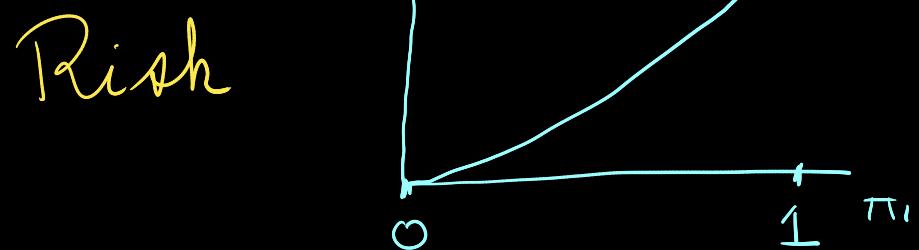
Return is a R.V. :

$$R = \pi_1 R_1 + (1 - \pi_1) R_2$$

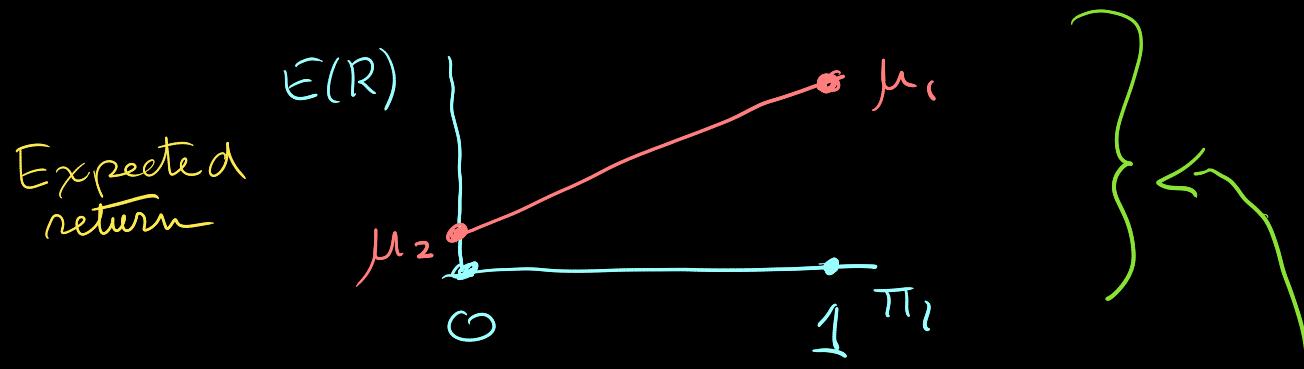
$$E(R) = \pi_1 \mu_1 + (1 - \pi_1) \mu_2$$



$$\text{Var}(R) = \pi_1^2 \sigma_1^2$$

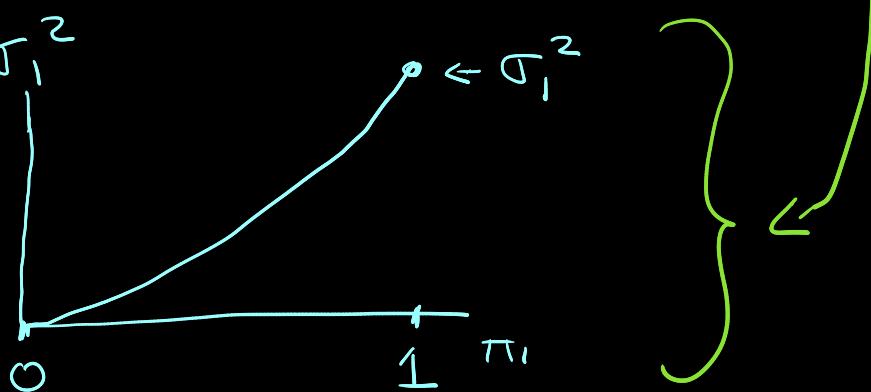


$$E(R) = \pi_1 \mu_1 + (1 - \pi_1) \mu_2$$



$$\text{Var}(R) = \pi_1^2 \sigma_1^2$$

Risk



Need to balance
expected return
against
risk.

- When $\sigma_2 > 0$ we will also need to consider covariance
- Variance is not the whole story as far as risk is concerned.

Read 4.2.1 Measurement error

Suppose we want to measure some "fixed" but "unknown" quantity.

Call it " x_0 ".

Measurement $X = x_0 + \beta + \varepsilon$

\uparrow \uparrow ↙

fixed systematic random
unknown error error

- same each
time

β = "bias"

ε = random component: $\text{Var}(\varepsilon) = \sigma^2$

$E(\varepsilon) = 0$

varies
with
each
measurement.

$$\text{Then } \text{Var}(X) = \text{Var}(\underbrace{\alpha_0 + \beta_0 + \varepsilon}_{\text{constant}})$$

$$= \text{Var}(\varepsilon) = \sigma^2$$

$$\text{But } E(\text{Error}^2) = \text{MSE}$$

$$= E[(X - \alpha_0)^2]$$

$$= E[(\beta_0 + \varepsilon)^2]$$

$$= E[\beta_0^2 + 2\beta_0 \varepsilon + \varepsilon^2]$$

$$= \beta_0^2 + 2\beta_0 E(\varepsilon) + E(\varepsilon^2)$$

$$= \beta_0^2 + 2\beta_0 \underbrace{E(\varepsilon)}_{=0} + E(\varepsilon^2)$$

$$= \text{Var}(X) + (E(X))^2 = \text{Var}(X) = \sigma^2$$

$$\text{"MSE"} = \beta_0^2 + \sigma^2 = \text{Bias}^2 + \sigma^2$$

4.3 Covariance & Correlation

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

"with"

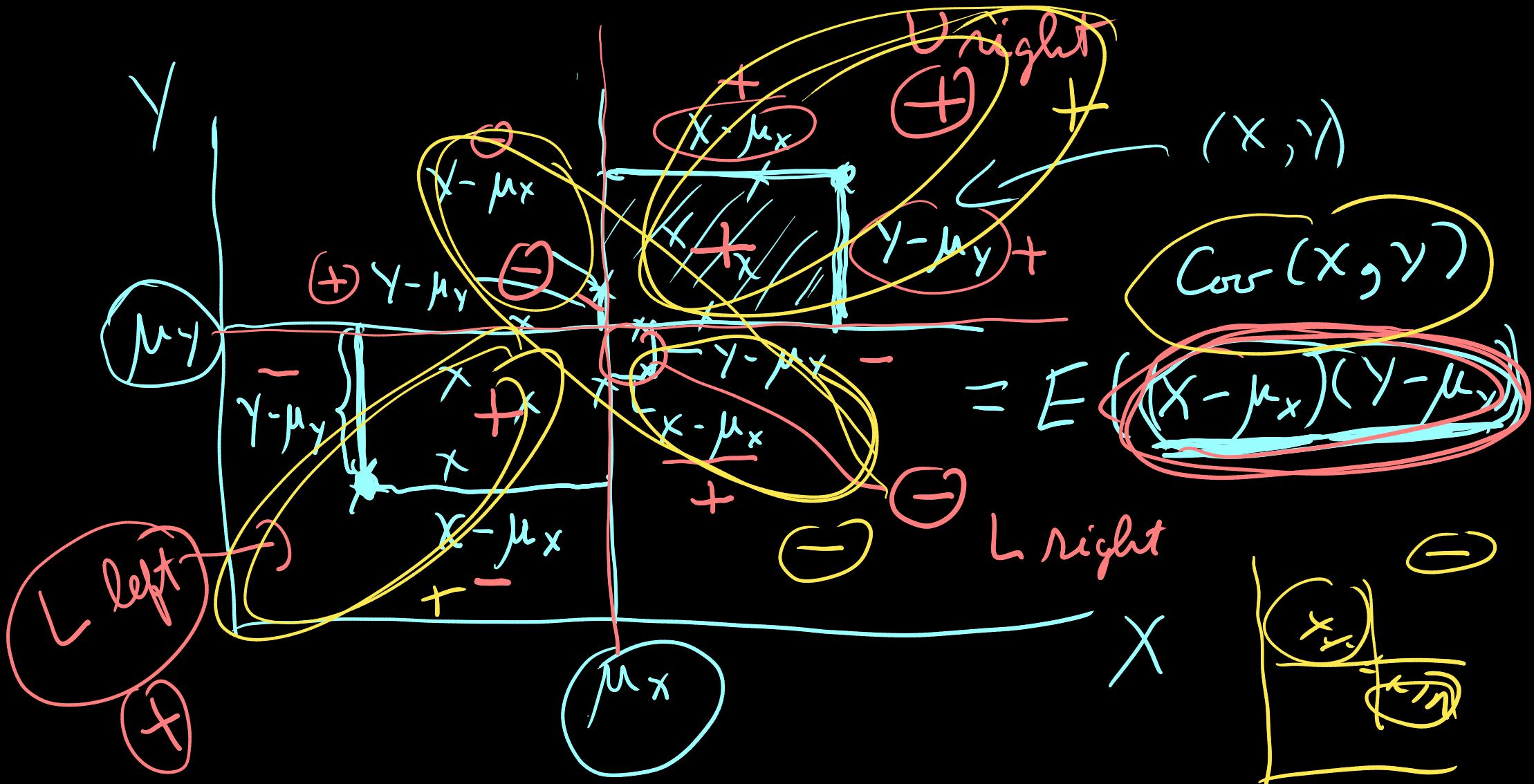
if E exists. "times"
is correct here

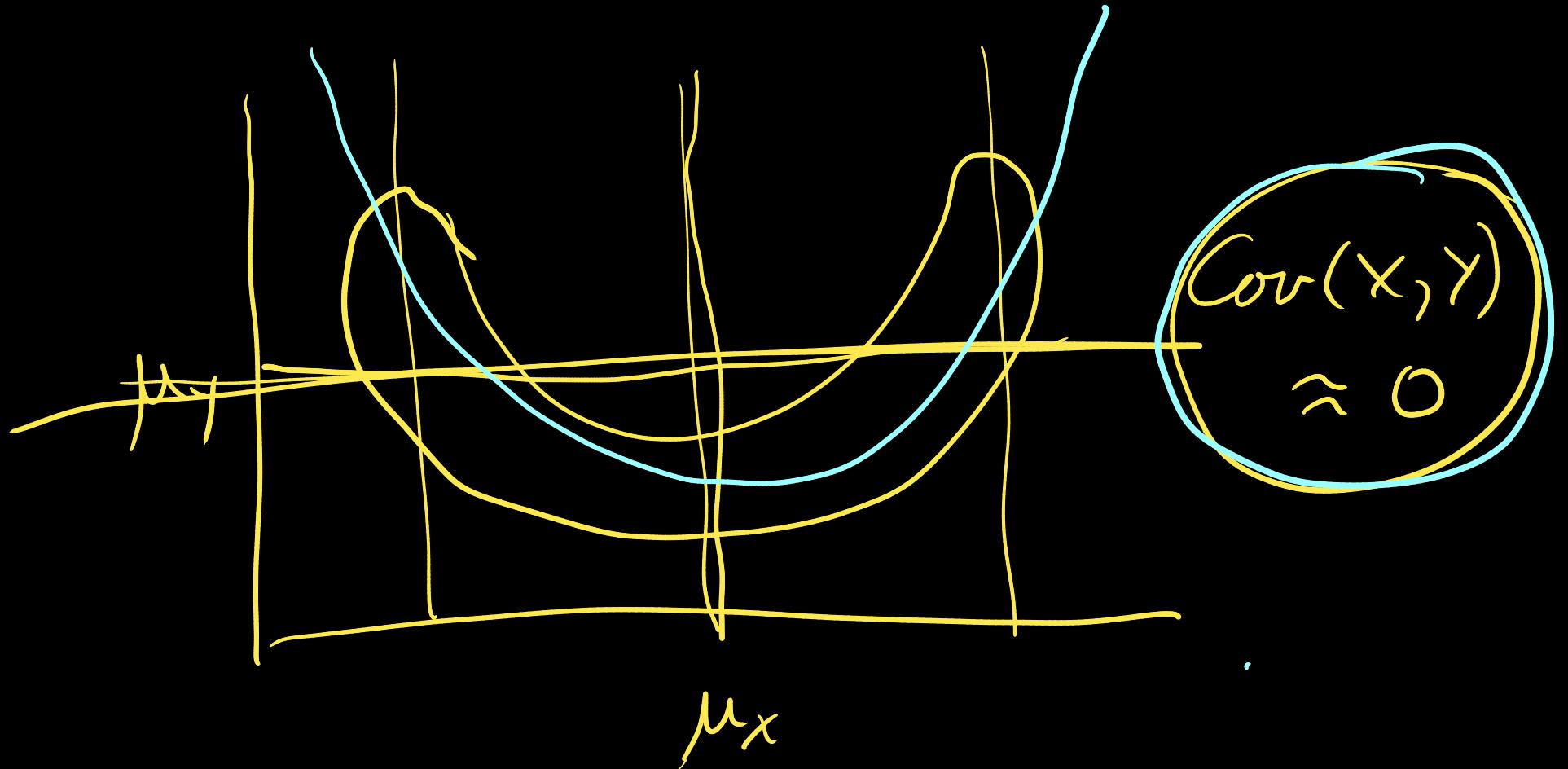
Fact: $\text{Cov}(X, Y) = E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]$

$$= E(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y$$

$$= E(XY) - E(X)E(Y)$$

never say
covariance of
~~X times Y~~
BUT always X "with" Y





(X, Y) joint Normal.

$$X \perp Y \quad \text{Cov}(X, Y) = 0$$

if

~~X & Y are independent~~

$$\text{Note : } \text{Cov}(X, X) \stackrel{?}{=} E((X - \mu_X)(X - \mu_X)) \\ = \text{Var}(X)$$

Facts

$$\text{Cov}(a + X, b + Y) = \text{Cov}(X, Y)$$

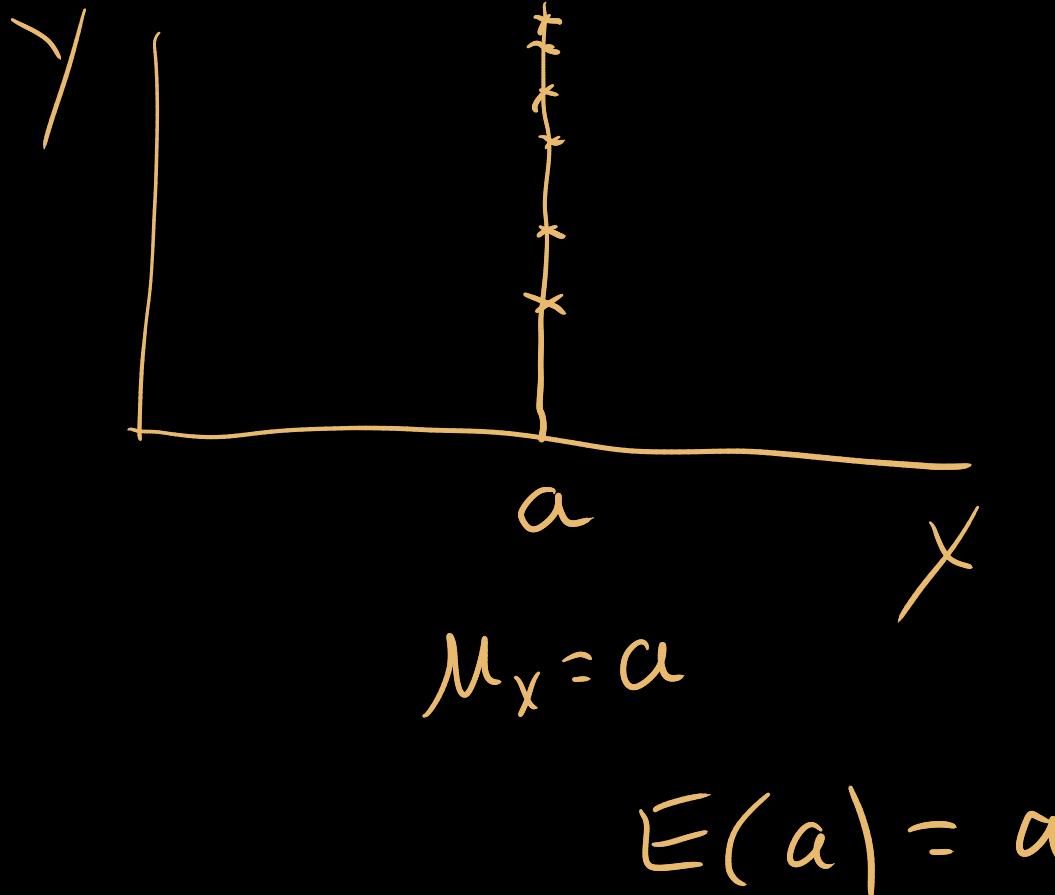
$$\text{Cov}(a\underline{X}, b\underline{Y}) = \underline{ab} \text{Cov}(X, Y)$$

$$\text{Cov}(\underline{X}, \underline{Y} + \underline{Z}) \\ = \underline{\text{Cov}(X, Y)} + \underline{\text{Cov}(X, Z)}$$

$$\text{Cov}(\underline{(X + W)}, \underline{Y}) = \underline{\text{Cov}(X, Y)} + \underline{\text{Cov}(W, Y)}$$

a, b
constants

$$\text{Cov}(a, y) = a \underbrace{\text{Cov}(y)}_{\text{not defined}}$$



$$\text{Cov}(X, Y)$$

with

$$\text{Cov}(x, x) = \text{Var}(x)$$

$$E((\underline{x - \mu_x})(\underline{y - \mu_y})) = 0$$



$$\text{Cov}(a + b_1 X_1 + b_2 X_2, Y) = \text{Cov}(a, Y) + [b_1 \text{Cov}(X_1, Y) + b_2 \text{Cov}(X_2, Y)]$$

linear in left argument
" " " right "

bilinear

$$\begin{aligned} \text{Cov}(aW + bX, cY + dZ) &= ac \text{Cov}(W, Y) + ad \text{Cov}(W, Z) \\ &\quad + bc \text{Cov}(X, Y) + bd \text{Cov}(X, Z) \end{aligned}$$

with

Summary:

Linearity of Expectation :

If $y = a + b_1x_1 + \dots + b_nx_n$

then

$$E(y) = a + b_1E(x_1) + \dots + b_nE(x_n)$$

In other words

If $y = a + \sum_{i=1}^n x_i$

then $E(y) = a + \sum_{i=1}^n E(x_i)$

Bilinearity of Covariance

$$\text{Cov}(X, a + \sum_{i=1}^n b_i Y_i)$$

$$= \sum_{i=1}^n b_i \text{Cov}(X, Y_i)$$

$$\text{Cov}(a + \sum_{i=1}^n b_i X_i, Y)$$

$$= \sum_{i=1}^n b_i \text{Cov}(X_i, Y)$$

$$\text{Cov}(a, Y) = 0$$

$$\text{Cov}(a + \sum_{i=1}^n b_i X_i, c + \sum_{j=1}^m d_j Y_j)$$
$$= \sum_{i=1}^n \sum_{j=1}^m b_i d_j \text{Cov}(X_i, Y_j)$$

*this should look
like matrix multiplication*

$$= [b_1 \ b_2 \ \dots \ b_n] x$$

m Y col.

$$\begin{matrix} \text{Cov}(x_1, y_1) & \text{Cov}(x_1, y_2) & \dots & \text{Cov}(x_1, y_m) \\ \text{Cov}(x_2, y_1) & \text{Cov}(x_2, y_2) & & \\ \vdots & \vdots & & \\ \text{Cov}(x_n, y_1) & \text{Cov}(x_n, y_2) & \dots & \text{Cov}(x_n, y_m) \end{matrix}$$

d_1

d_2

\vdots

d_m

$$(l \times m) \times (n \times m) \times (m \times 1) = l \times 1$$

vector matrix vector

$$[\quad] \text{ scalar}$$

Conformable

Matrix formula:

Let

$$\tilde{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$$C = \begin{bmatrix} \text{Cov}(X_i, Y_i) \end{bmatrix}$$

$n \times m$ matrix

$$\tilde{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix}$$

$$\text{Cov}(\tilde{b}^T X, \tilde{d}^T Y)$$

$$= \tilde{b}^T C \tilde{d}$$

$$(b_1, \dots, b_m)(x_1, \dots, x_n)$$

\underline{b}) and $\underline{d})$ could be matrices

$$\text{Cov}(\underline{B}\underline{x}, \underline{D}\underline{y}) = \begin{bmatrix} \underline{B} \text{Cov}(\underline{x}, \underline{y}) \underline{D}^T \end{bmatrix} = \underline{b} \cdot \underline{c}$$

Fact:

If X & Y are independent

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

Proof: Recall: if X, Y are independent then $E(g(X)h(Y))$
 $= \underline{E(g(X))} \underline{E(h(Y))}$

Let $g(x) = \underline{x - \mu_x}$, $h(y) = \underline{y - \mu_y}$

$$\text{Cov}(x, y) = E[(x - \mu_x)(y - \mu_y)]$$

$$= E(x - \mu_x) \otimes E(y - \mu_y)$$

$$= [\frac{\mu_x - \mu_x}{0}] \times [\frac{\mu_y - \mu_y}{0}]$$

$$= \underline{0}$$

Important to remember :

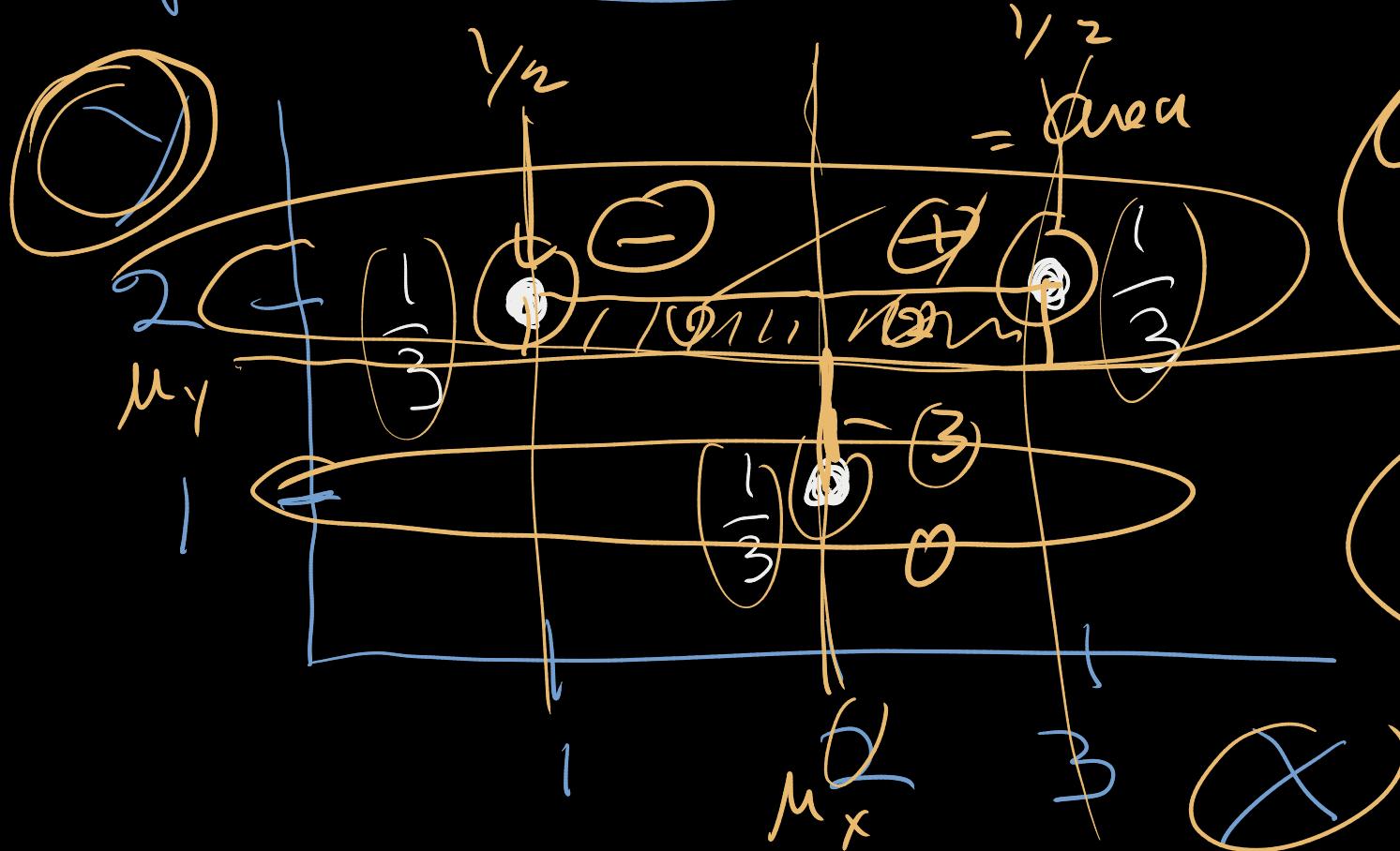
If (X, Y) are bivariate normal
then X and Y are independent
if $\text{Cov}(X, Y) = 0$

In general: X, Y independent

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

BUT $\text{Cov}(X, Y) = 0$ ~~DOES NOT~~ IMPLIES INDEPENDENCE

Counter Simple Example



IMPORTANT RESULT

If (X_1, \dots, X_n) are R.V.

with $\text{Cov}(X_i, X_j) = 0$ if $i \neq j$

then

$$\text{Var}(X_1 + \dots + X_n)$$

$$= \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

Corollary : If X_1, \dots, X_n

are independent, then

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

NOTE: NOT TRUE IN GENERAL

Recall :

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

whether X_i 's are independent or not

Example B (p. 140)

$$X \sim \text{Binomial}(n, p)$$

What is $\text{Var}(X)$?

$$X = X_1 + \dots + X_n$$

where X_i 's are

independent Bernoulli (p)

$$E(X_i) = p \cdot 1 + (1-p) \cdot 0 = p$$

$$E(X_i^2) = p \cdot 1^2 + (1-p) \cdot 0^2 =$$

$$\begin{aligned}\text{Var}(X_i) &= E(X^2) - (E(X))^2 \\&= p - p^2 \\&= p(1-p) \\&= pq \quad \text{where } q = 1-p\end{aligned}$$

$$\text{So } \text{Var}(X) = \sum_{i=1}^n (\text{Var}(X_i)) = npq$$

Correlation:

"Pure" measure of covariance

X = height in cm

Y = weight in kg

then $\text{Cov}(X, Y)$ is in "cm \times kg"

If we change X to $X' = X / 2.54$

so X' is in inches, then

$$\text{Cov}(X', Y) = \text{Cov}(X, Y) / 2.54$$

in units of "in \times kg"

$$\begin{aligned}\text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{SD(X) \times SD(Y)} \\ &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \times \text{Var}(Y)}} \\ &= \rho_{XY} = \rho\end{aligned}$$

Theorem : $|ρ| \leq 1$ i.e. $-1 \leq ρ \leq 1$
and $ρ = \pm 1$ iff $P(Y = a + bX) = 1$
for some $b \neq 0$

See proof p. 143 = Cauchy-Schwarz Theorem

Fact: $ρ_{XY}$ stays the same if X and/or Y
are rescaled or translated.

e.g. inches to feet, ${}^{\circ}\text{C} \rightarrow {}^{\circ}\text{F}$
 $X' = X/12$ $X' = 9/5 X + 32$

