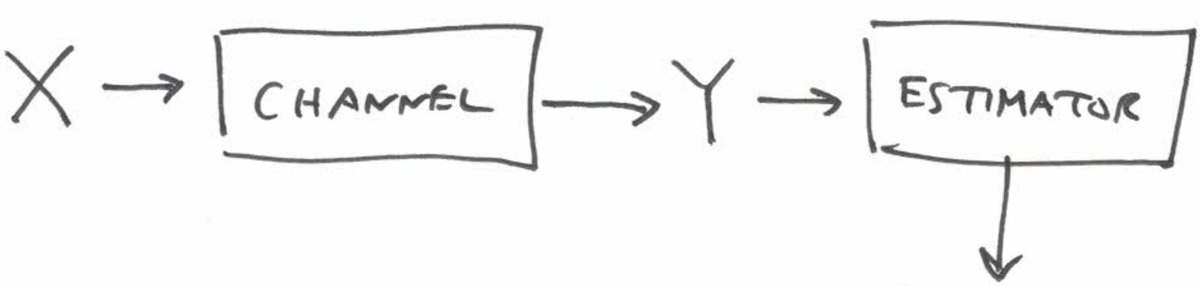
ESTIMATION

SETUP: WE <u>OBSERVE</u> THE VALUE OF A R.V. Y AND WE WANT TO <u>ESTIMATE</u> THE VALUE OF A RELATED R.V. X. (WE DON'T OBSERVE X DIRECTLY).

WE HAVE INFO LIKE CONDITIONAL/ JOINT PDFS.



$$\hat{X} = g(Y)$$

WE WANT TO FIND $g(Y)$ SO THAT $g(Y)$ IS "AS CLOSE AS POSSIBLE TO" OR "THE BEST ESTIMATE OF" X.

MAP ESTIMATION

MAXIMUM A POSTERIORI

"AFTER THE FACT"

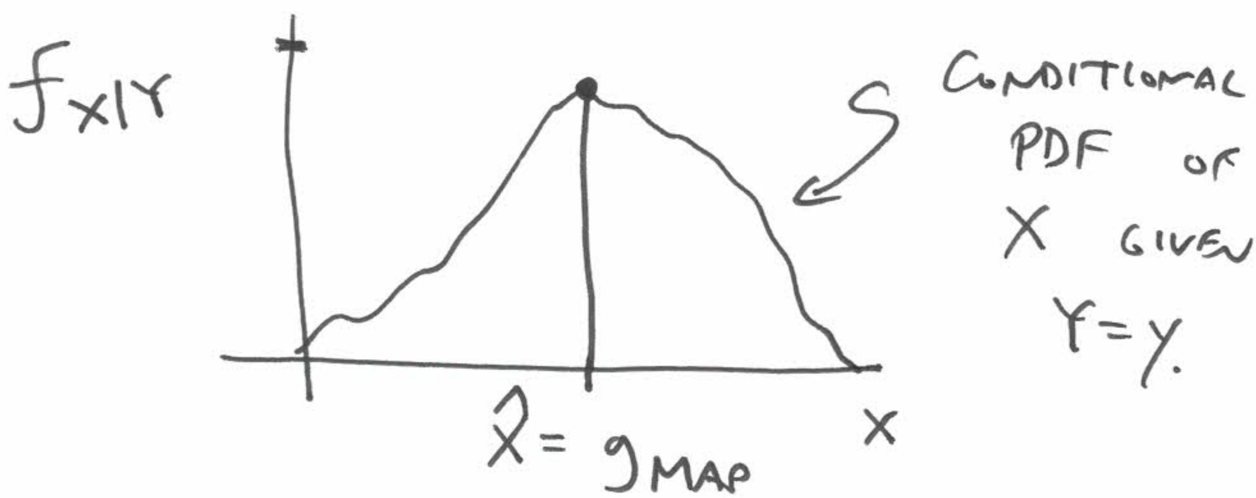
GIVEN THE VALUE OF Y,

WHAT VALUE OF X IS

MOST LIKELY TO HAVE OCCURRED?

$$\hat{X} = g_{MAP}(Y)$$

$$= \arg\max_{x} f_{X|Y}(x|y)$$



RECALL: FLIP COIN 3 TIMES

X = # HEADS

Y = POS FIRST HEAD

$X \backslash Y$	0	1	2	3
0	$\frac{1}{8}$			
1		$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2		$\frac{1}{4}$	$\frac{1}{8}$	
3		$\frac{1}{8}$		
	$1/8$	$1/2$	$1/4$	$1/8$

JOINT PMF

$f_{X\|Y}$	0	1	2	3
0	1	0	0	0
1	0	$\frac{1}{4}$	$\frac{1}{2}$	1
2	0	$\frac{1}{2}$	$\frac{1}{2}$	0
3	0	$\frac{1}{4}$	0	0

($Y=2$ CAN CHOOSE $X = 1$ OR 2)

$$g_{MAP}(Y) = \begin{cases} 0 & Y=0 \\ 2 & Y=1 \\ 1 \text{ or } 2 & Y=2 \\ 1 & Y=3 \end{cases}$$

PROBABILITY OF ERROR:

$$P_E = \sum_{k=0}^{3} P(\text{ERROR} \mid Y=k)\, P(Y=k)$$

$$= 0 \cdot \frac{1}{8} + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} + 0 \cdot \frac{1}{8}$$

$$= \frac{3}{8}$$

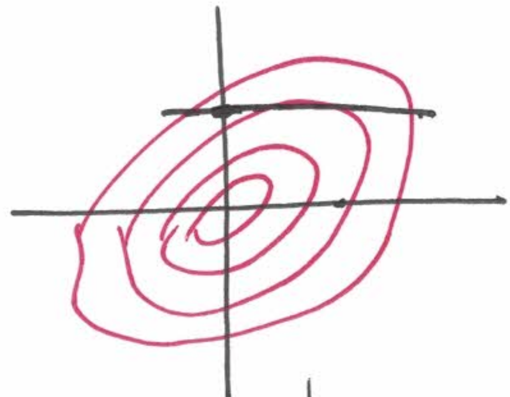
OR JUST ADD UP ENTRIES OF JOINT PMF.

EX X AND Y ARE JOINTLY

GAUSSIAN, $\mu_X = \mu_Y = 0$

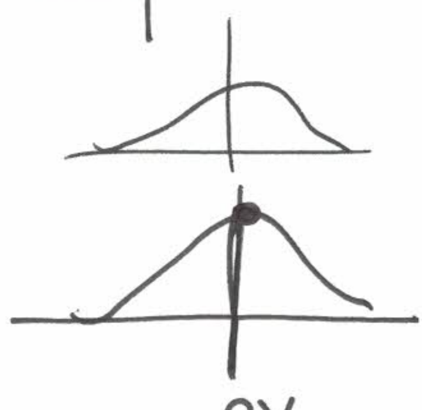
$\sigma_X = \sigma_Y = 1$

ρ = CORR. COEFF $\in [-1, 1]$



WE SHOWED THAT

$f_{X|Y}(x|y)$



WAS ALSO GAUSSIAN

WITH MEAN ρy

VARIANCE $\sqrt{1-\rho^2}$.

SO $\hat{X}_{MAP} = g_{MAP}(y) = \rho y$.

IN REAL LIFE, WE MAY NOT KNOW / BE ABLE TO ESTIMATE $f_{X|Y}(x|y)$.

IT'S OFTEN MUCH EASIER TO OBTAIN $f_{Y|X}(y|x)$.

e.g., PUT A LOT OF X VALUES INTO THE CHANNEL, BUILD $f_{Y|X}$.

KNOWN DURING TEST

OBSERVED

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)}$$

OBSERVE.

MAXIMUM LIKELIHOOD ESTIMATION

GIVEN THE VALUE OF Y, WHAT VALUE OF X IS MOST LIKELY TO HAVE PRODUCED IT?

$$\hat{X} = g_{ML}(Y) = \arg\max_{x} f_{Y|X}(y|x)$$

DISTINCTION B/W MAP AND ML:

MAP ESTIMATE TAKES INTO ACCOUNT UNDERLYING PROBABILITY OF X. ML DOESN'T TAKE THIS INTO ACCOUNT.
"UNINFORMATIVE PRIOR"

COIN FLIP EXAMPLE:

$f_{Y\mid X}$	$Y=0$	1	2	3
$X=0$	1	0	0	0
1	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
2	0	$\frac{2}{3}$	$\frac{1}{3}$	0
3	0	1	0	0

$$g_{ML}(Y) = \begin{cases} 0 & Y=0 \\ 3 & Y=1 \\ 2 & Y=2 \\ 1. & Y=3 \end{cases}$$

P_{ERR} IN THIS CASE:

$$0 \cdot \frac{1}{8} + \frac{3}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} + 0 \cdot \frac{1}{8}$$

$$= \frac{4}{8} = \frac{1}{2} > \frac{3}{8}$$ FROM MAP ESTIMATE.

ML EXAMPLE FOR GAUSSIAN:

$f_{Y|X}$ GAUSSIAN MEAN ρx

VARIANCE $\sqrt{1-\rho^2}$

$$f_{Y|X} = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}$$

$$\frac{d}{dx} = 0.$$

$$2(y-\rho x) \cdot -\rho \; e^{-(\cdot\ \cdot)}$$

$$\downarrow$$

$$y - \rho x = 0 \qquad \hat{x}_{ML.} = \frac{y}{\rho}.$$

SO $\quad g_{ML}(y) = \dfrac{y}{\rho} \neq g_{MAP}(y) = y\rho.$

DIFFERENT DEFINITION OF "BEST":

MINIMUM MEAN-SQUARE ESTIMATION: (MMSE)

$$\min \; E\left(\left(g(Y) - X\right)^2 \right)$$

i.e. MINIMIZE DEVIATION BETWEEN X AND $g(Y) = \hat{X}$ OVER ALL POSSIBLE FUNCTIONS $g(Y)$.

EASIER PROBLEM: WHAT IS THE CONSTANT c THAT MINIMIZES

$$E\left((c - X)^2\right) \; ?$$

$$\min_{c} \; E\left((c-x)^2 \right)$$

$$= E\left(c^2 - 2cX + X^2 \right)$$

$$= \cancel{E}\, c^2 - 2cE(x) + E(x^2)$$

$$\frac{d}{dc} = 0 \Rightarrow 2c - 2E(x) = 0.$$

$$c = E(x).$$

MAKES SENSE; MEAN VALUE
MINIMIZES THE ERROR.

NOW CONSIDER

$$E\left(\left(g(Y)-X\right)^2\right)$$

$$= E\left(E\left(\left(g(Y)-X\right)^2 \mid Y=y\right)\right)$$

[LAW OF ITERATED EXPECTATION]

$$\int_{-\infty}^{\infty} E\left(\left(g(y)-X\right)^2 \mid Y=y\right) f_Y(y)\,dy$$

THIS INTEGRAL WILL BE MINIMIZED IF WE MINIMIZE

$$E\left(\left(g(Y)-X\right)^2 \mid Y=y\right) \text{ FOR}$$

EACH POSSIBLE VALUE OF y.

FOR FIXED $Y=y$, $g(Y)$ IS JUST A CONSTANT.

FROM PREVIOUS RESULT, THIS CONSTANT SHOULD BE

$$E(X \mid Y=y)$$

$$g_{MMSE}(Y) = E(X \mid Y)$$

IN PRACTICE, THIS COULD BE MESSY (SOME WEIRD FUNCTION OF Y).

COMPROMISE: TAKE THE BEST LINEAR FUNCTION OF Y, $g(Y) = aY + b$, TO MINIMIZE THE MEAN-SQUARED ERROR

$$\min_{a,b} E\left(\left((aY + b) - X\right)^2\right)$$

$$E\left(\big(\underbrace{b}_{\text{CONSTANT}} - (X - aY)\big)^2\right)$$

$b^* = E(X - aY)$ FROM BEFORE

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

$$a^* = \arg\min_{a} E\left(\left(E(X) - aE(Y) - X + aY\right)^2\right)$$

$$= E\left(\left(a(Y - E(Y)) - (X - E(X))\right)^2\right)$$

$$= E\Big(a^2 (Y - E(Y))^2$$
$$- 2a(Y - E(Y))(X - E(X))$$
$$+ (X - E(X))^2 \Big)$$

$$= a^2 \operatorname{Var}(Y) - 2a \operatorname{Cov}(X,Y) + \operatorname{Var}(X)$$

$$\frac{d}{da} = 0 \Rightarrow 2a\operatorname{Var}(Y) - 2\operatorname{Cov}(X,Y) = 0.$$

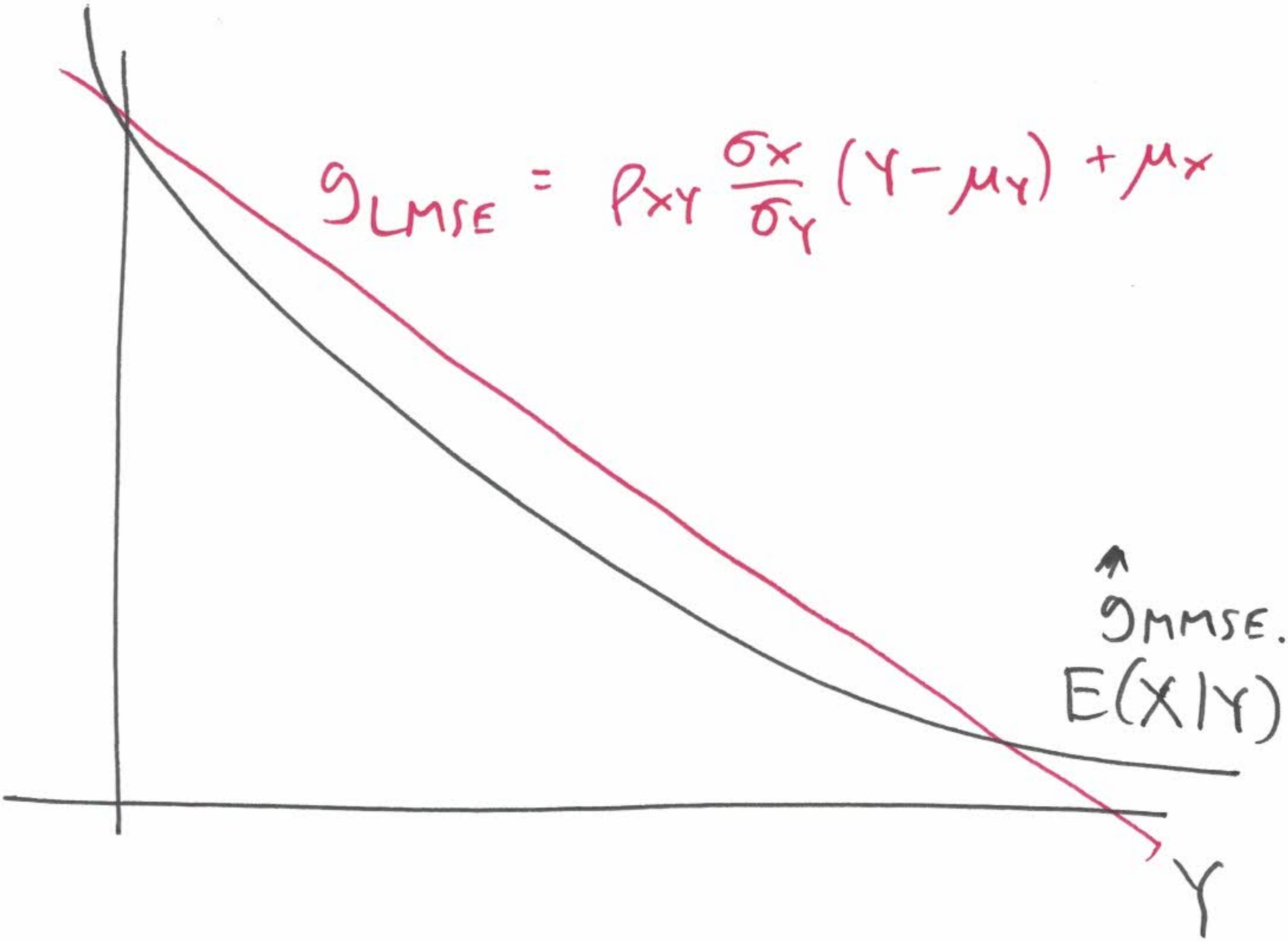
$$a^* = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)} = \frac{\rho_{XY}\sigma_X\sigma_Y}{\sigma_Y^2} = \rho_{XY}\frac{\sigma_X}{\sigma_Y}$$

THUS THE MMSE <u>LINEAR</u> ESTIMATOR FOR X IS:

$$\hat{X} = g_{LMSE}(Y)$$

$$= \rho_{XY}\frac{\sigma_X}{\sigma_Y}Y + \left(\mu_X - \rho_{XY}\frac{\sigma_X}{\sigma_Y}\mu_Y\right)$$

$$= \rho_{XY}\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y) + \mu_X$$



$g_{LMSE} = \rho_{XY} \frac{\sigma_X}{\sigma_Y}(Y - \mu_Y) + \mu_X$
$\hat{g}_{MMSE}$
$E(X|Y)$
Y