Estimation: LLSE

Aug 8, 2022

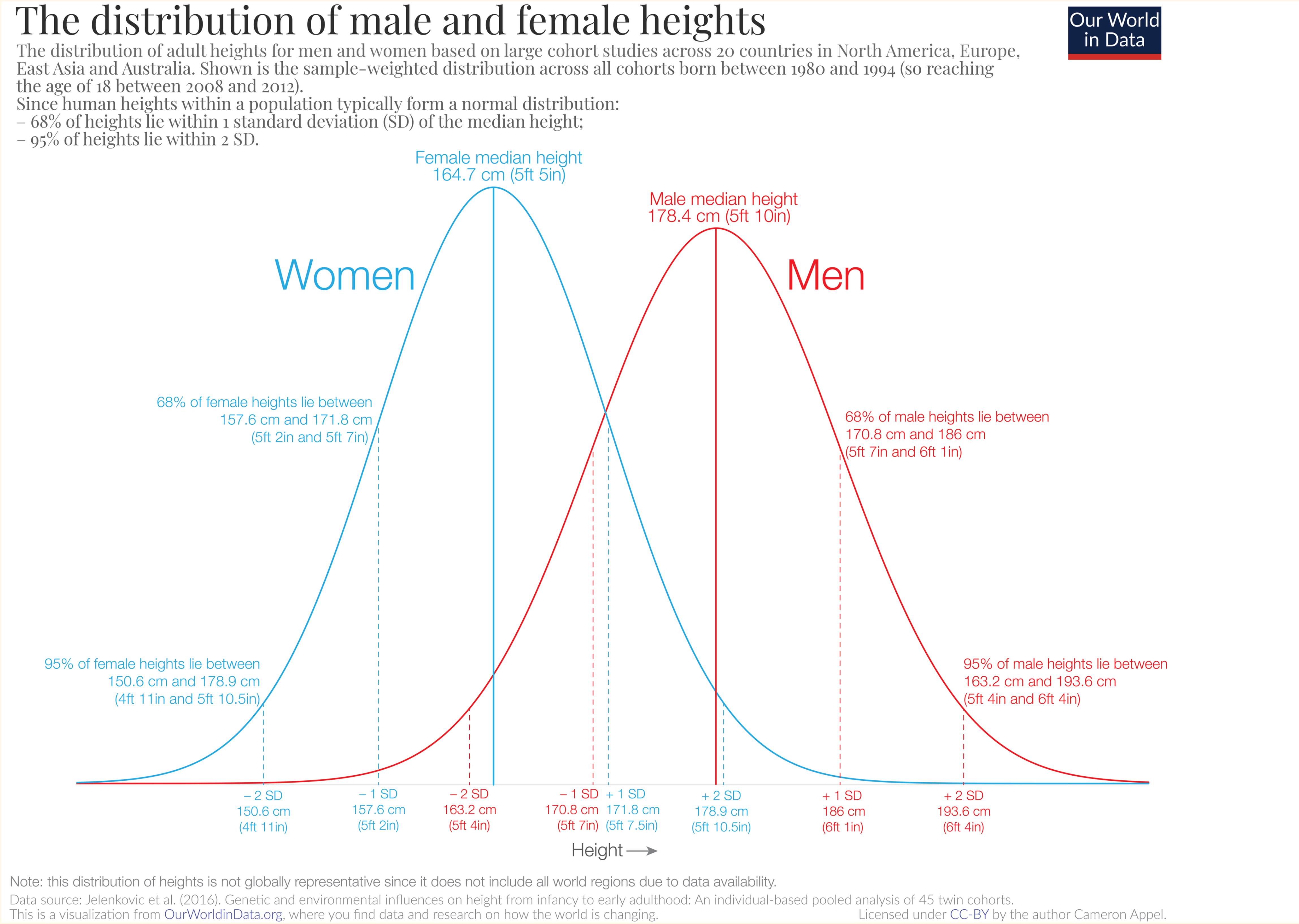
Uniform distribution

- $X \sim U[a,b]$
- $E(X) = (b+a)/2$
- $E(X^2) = (a^2+ab+b^2)/3$
- $Var(X) = (b-a)^2/12$

- final logistics on piazza
- email cs70-staff@berkeley.edu for admin issue
- course eval (80%)

- DSP long exam
- cheat sheet : 2 lettersize hand written
- OH today
- schedule.

Height of the person who sits next to you



https://ourworldindata.org/human-height#distribution-of-adult-heights

Mean squared error (MSE)

- We want to estimate value of a random variable, in absence of observations. All we know is the distribution of Y. Find a good estimator.
- How good an estimator is?

$$MSE = \mathbb{E}((Y - \hat{y})^2)$$

- The optimal estimator of Y is the one minimizes MSE.

$$MSE = \mathbb{E}(Y^2 + \hat{y}^2 - 2\hat{y}Y) = \mathbb{E}(Y^2) + \hat{y}^2 - 2\hat{y}\,\mathbb{E}(Y)$$

$$= Var(Y) + \mathbb{E}(Y)^2 + \hat{y}^2 - 2\hat{y}\,\mathbb{E}(Y)$$

$$\frac{dMSE}{d\hat{y}} = 2\hat{y} - 2\mathbb{E}(Y) = 0$$

$$\Rightarrow \quad \hat{y} = \mathbb{E}(Y)$$

$$MSE = \mathbb{E}\left((Y - \mathbb{E}(Y))^2\right) = Var(Y)$$

Height of the person who sits next to you

Y

estimator $g(X)$

- Now we have observation of this person's weight. X

$$MSE = \mathbb{E}((Y - g(X))^2 \mid X) = \mathbb{E}(Y^2 + g(X)^2 - 2g(X)Y \mid X)$$

$$= var(Y^2 \mid X) + \mathbb{E}(Y \mid X)^2 + g(X)^2 - 2g(X)\mathbb{E}(Y \mid X)$$

$$\frac{dMSE}{d(g(X))} = 2g(X) - 2\mathbb{E}(Y \mid X) = 0$$

$$g(X) = \mathbb{E}(Y \mid X)$$

Mean squared estimator

- A pair (X,Y) of random variables with joint distribution
- Generally, the mean squared estimation error associated with an estimator $g(X)$ is defined as

MSE = $$\mathbb{E}\left(\left(Y - g(X)\right)^2\right)$$

- $\mathbb{E}\left(\left(Y - g(X)\right)^2\right)$ is minimized when $g(X) = \mathbb{E}(Y|X)$.

Example 1.

Let Y be uniformly distributed over the interval [4, 10] and suppose that we observe X with some random error W. In particular, we observe the value of random variable

$$X = Y + W$$

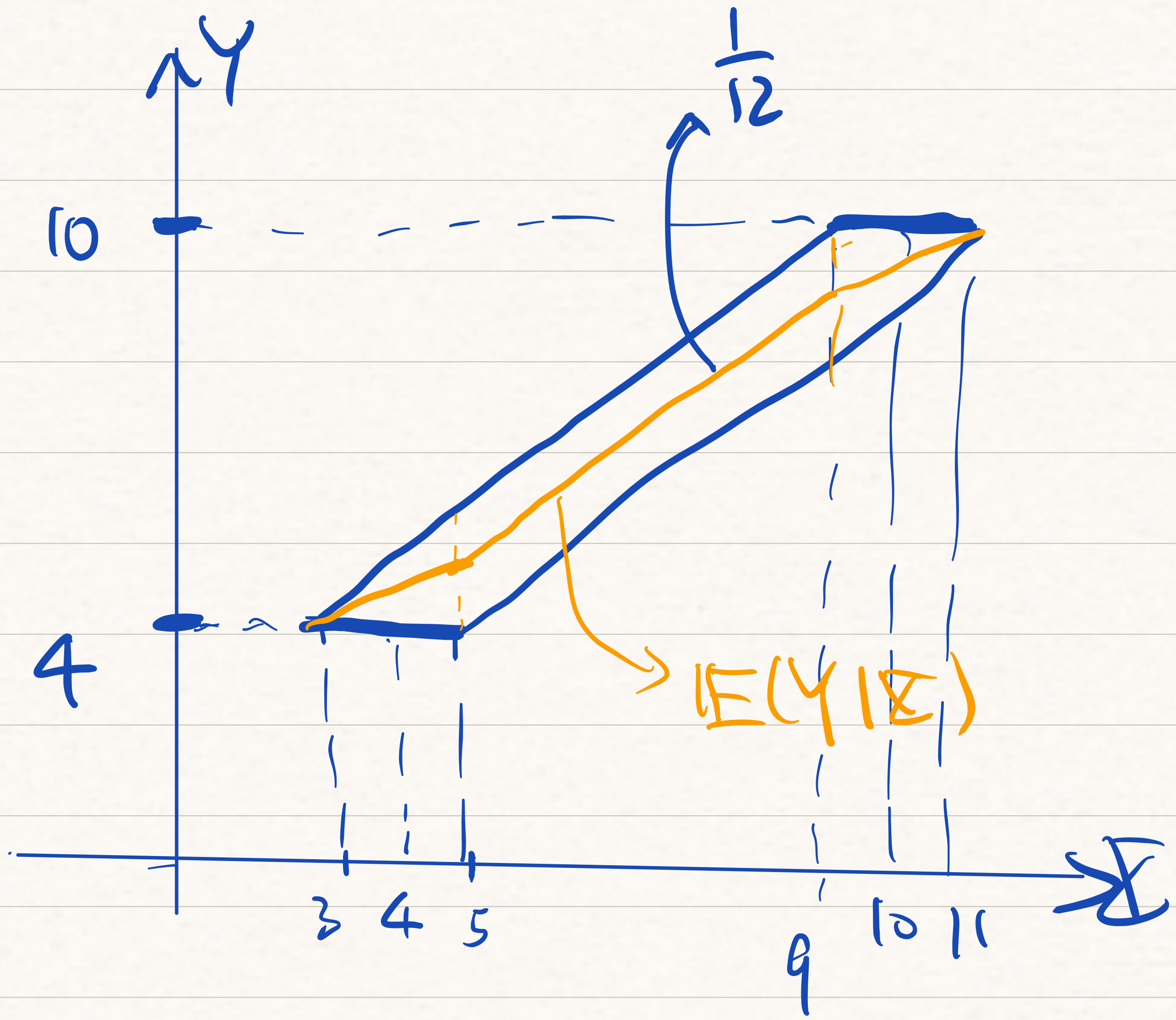
Assume that noise W is uniformly distributed over interval [-1, 1] and independent of Y.

$Y \sim U[4,10]$. $f_Y(y) = \begin{cases} \frac{1}{6} & y \in [4,10] \\ 0 & \text{o.w.} \end{cases}$ ←

condition on $Y=y$ $X = y + W \sim U[y-1, y+1]$ ←

joint PDF of X, Y is

$$f_{XY}(x,y) = f_{X|Y}(x|y)\, f_Y(y) = \begin{cases} \frac{1}{6}\cdot\frac{1}{2} = \frac{1}{12} & \text{if } y \in [4,10] \text{ and } x \in [y-1, y+1] \\ 0 & \text{o.w.} \end{cases}$$



$\hat{y} = g(X)$ is $\mathbb{E}(Y|X)$

$x \in (5,9)$, $Y|X \sim U[x-1, x+1]$

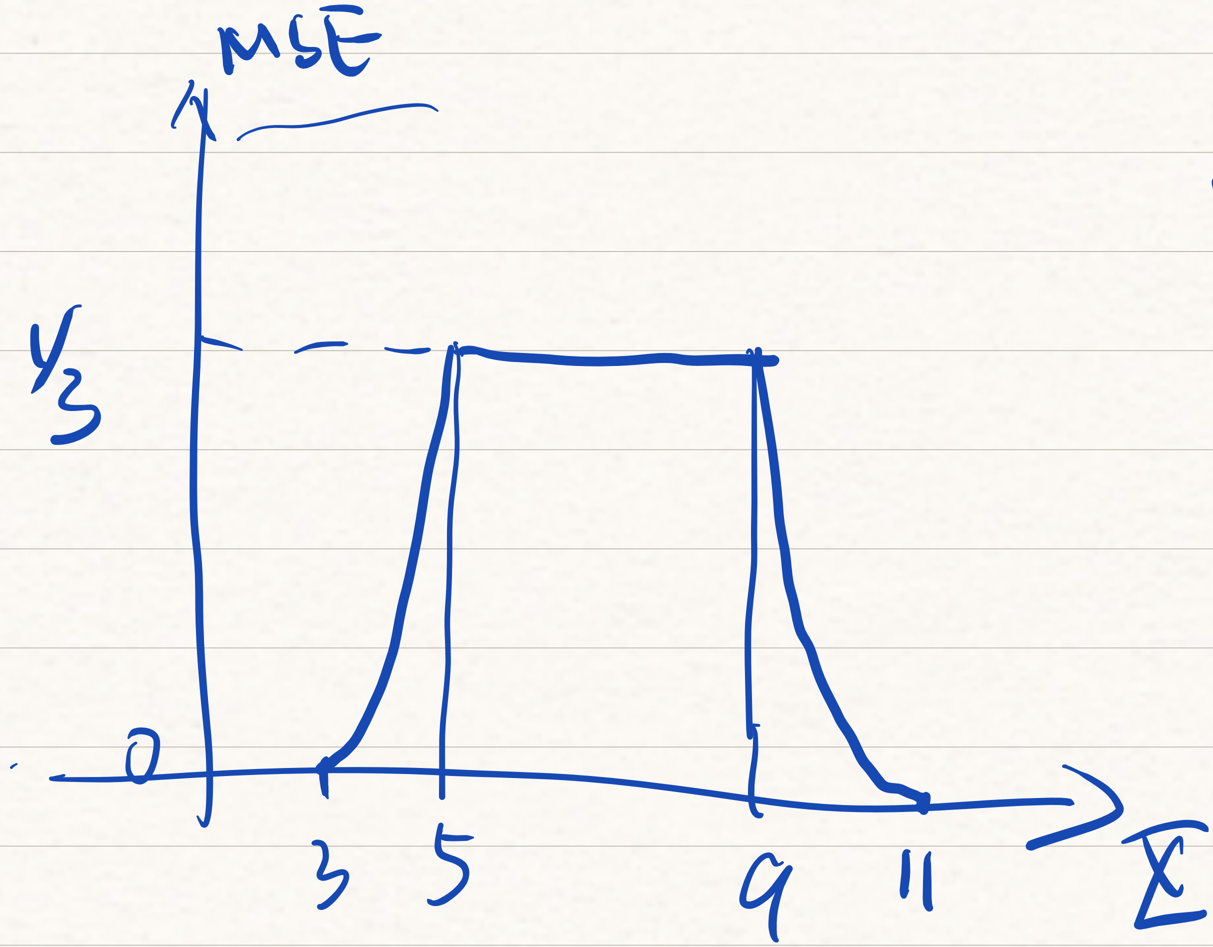
$x \in (3,5)$, $Y|X \sim U[4, x+1]$ ←

$x \in (9,11)$, $Y|X \sim U[x-1, 10]$

$\underline{\mathbb{E}((Y - \mathbb{E}(Y))^2 | X)} = Var(Y|X)$

MSE

$$Var(U(a,b)) = \frac{(b-a)^2}{12}$$



$$\mathrm{Var}(Y|X) = \begin{cases} \frac{(x+1-(x-1))^2}{12} = \frac{1}{3} & x \in (5,9) \\ \frac{(x+1-4)^2}{12} = \frac{(x-3)^2}{12} & x \in (3,5) \\ \frac{(10-(x-1))^2}{12} = \frac{(11-x)^2}{12} & x \in (9,11) \\ 0 & \text{o.w.} \end{cases}$$

for $X=3$, for sure we know $Y=4$

Linear least squared estimation

The LLSE of Y given X, denoted by $L[Y|X]$*, is the linear function* $g(X) = a + bX$ *that minimizes*

$$\hat{y}$$

$$MSE = \mathbb{E}((Y - a - bX)^2) \leftarrow$$

$\hookrightarrow L[Y|X]$

$$MSE = \mathbb{E}(Y^2 + a^2 + b^2X^2 - 2aY - 2bYX + 2abX)$$

$$= \mathbb{E}(Y^2) + a^2 + b^2\mathbb{E}(X^2) - 2a\mathbb{E}(Y) - 2b\mathbb{E}(XY) + 2ab\mathbb{E}(X)$$

$$\frac{dMSE}{da} = 2a - 2\mathbb{E}(Y) + 2b\mathbb{E}(X) = 0 \quad (1)$$

$$\frac{dMSE}{db} = 2b\mathbb{E}(X^2) - 2\mathbb{E}(XY) + 2a\mathbb{E}(X) = 0 \quad (2)$$

$$a = \mathbb{E}(Y) - b\,\mathbb{E}(X)$$

$$b\,\mathbb{E}(X^2) - \mathbb{E}(XY) + \mathbb{E}(Y)\mathbb{E}(X) - b\,\mathbb{E}(X)^2 = 0$$

$$b\big(\underbrace{\mathbb{E}(X^2) - \mathbb{E}(X)^2}_{var(X)}\big) - \big(\underbrace{\mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X)}_{cov(X,Y)}\big) = 0$$

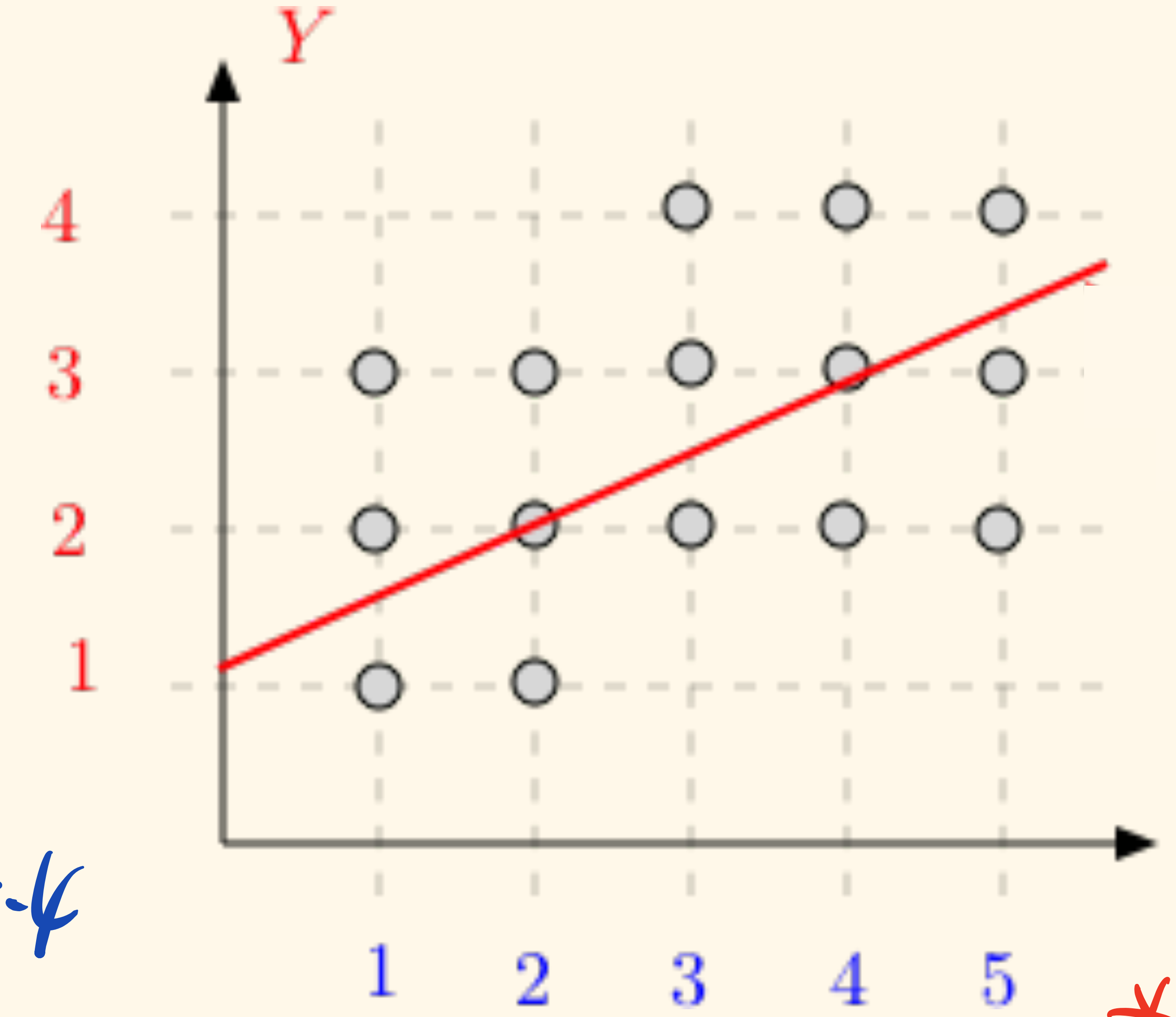
$$b = \frac{cov(X,Y)}{Var(X)} \qquad a = \mathbb{E}(Y) - \frac{cov(X,Y)}{Var(X)}\mathbb{E}(X)$$

$$\hat{y} = \mathcal{L}[Y|X] = g(X) = a + bX = \mathbb{E}(Y) - \frac{cov(X,Y)}{Var(X)}\mathbb{E}(X) + \frac{cov(X,Y)}{Var(X)}X$$

$$= \mathbb{E}(Y) + \frac{cov(X,Y)}{Var(X)}\big(X - \mathbb{E}(X)\big)$$

Example 2.

Consider discrete joint distribution of X and Y



$\mathbb{E}(X) = 3 \quad \mathbb{E}(Y) = 2.5$?

$\mathbb{E}(X^2) = \frac{3}{15}(1+2^2+3^2+4^2+5^2) = 11$

$\mathbb{E}(XY) = \frac{1}{15}(1\times 1, 1\times 2 + \cdots + 5\times 4) = 8.4$

$\mathrm{Var}(X) = 11 - 3^2 = 2$

$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0.9$

$\hat{Y} = L(Y \mid X) = 2.5 + \frac{0.9}{2}(X - 3)$

$= 1.15 + 0.45X$

$\mathbb{E}(XY) = \iint xy\, f_{XY}(x,y)\,dx\,dy$ **

$= \sum_x \sum_y xy\, \mathbb{P}(X=x, Y=y)$

correction : $\mathbb{E}(X) = 3.$

$$\mathbb{E}(Y) = \frac{2}{15} + \frac{2}{3} + \frac{3}{3} + \frac{4}{5} = \frac{39}{15} = 2.6$$

$$\mathbb{E}(X^2) = \frac{3}{15}(1^2 + 2^2 + 3^2 + 4^2 + 5^2) = 11$$

$$\mathbb{E}(XY) = \frac{1}{15}(1\times1 + 2\times1 + 1\times2 + 2\times2 + 3\times2 + 4\times2 + 5\times2$$
$$+ 1\times3 + 2\times3 + 3\times3 + 4\times3 + 5\times3$$
$$+ 3\times4 + 4\times4 + 5\times4) = 8.4$$

$Var(X) = 11 - 9 = 2$ $\quad$ $cov(X,Y) = 8.4 - 7.8 = 0.6$

$$L(Y|X) = 2.6 + \frac{0.6}{2}(X - 3) = 0.3X + 1.1$$

Linear regression non bayesian.

For X and Y, We observe K samples $(X_1, Y_1) \ldots (X_K, Y_K)$. $\widehat{Y_n} = a + bX_n$ is the guess of Y_n given X_n.

We want to find the value a and b to minimize the mean squared error

$$MSE = \frac{1}{K}\sum_{k=1}^{K}(Y_k - a - bX_k)^2$$

Linear regression of Y over X is

$$\hat{Y} = a + bX = \mathbb{E}(Y) + \frac{cov(X,Y)}{var(X)}(X - \mathbb{E}(X))$$

Where $\mathbb{E}(Y) = \frac{1}{K}\sum_{k=1}^{K} Y_k$, $\mathbb{E}(X) = \frac{1}{K}\sum_{k=1}^{K} X_k$, ← sample mean

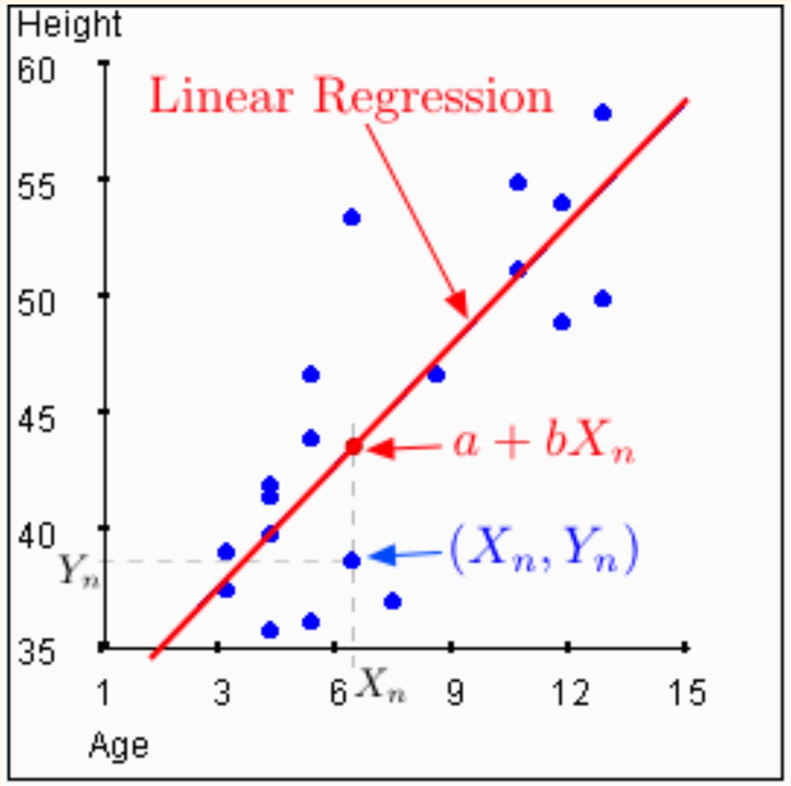
$var(X) = \frac{1}{K}\sum_{k=1}^{K}(X_k - \mathbb{E}(X))^2$, $cov(X,Y) = \frac{1}{K}\sum_{k=1}^{K} X_k Y_k - \mathbb{E}(X)\mathbb{E}(Y)$

sample variance sample covariance

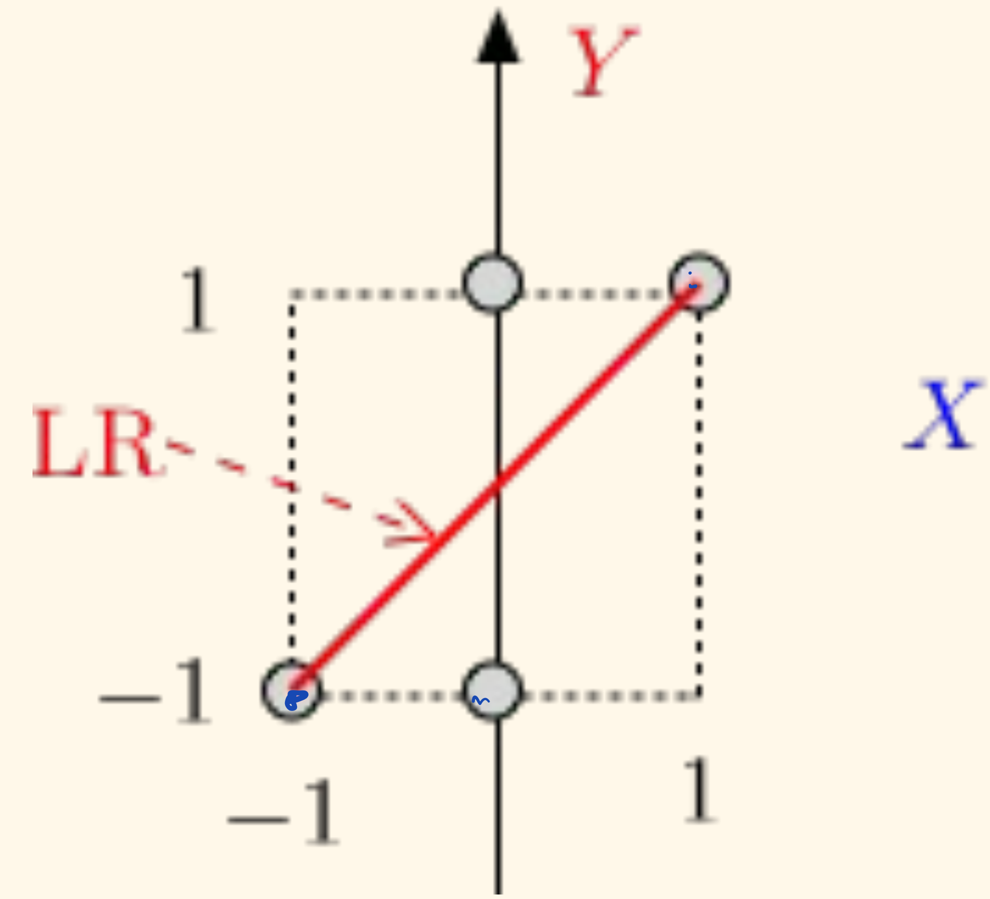
Linear regression converge to LLSE

For X and Y, We observe K samples $(X_1, Y_1) \ldots (X_K, Y_K)$. Assume that samples are i.i.d. As sample number increases, the linear regression approaches LLSE of X, Y.

by LLN



Example 3



$\mathbb{E}(X) = 0 \quad \mathbb{E}(Y) = 0$

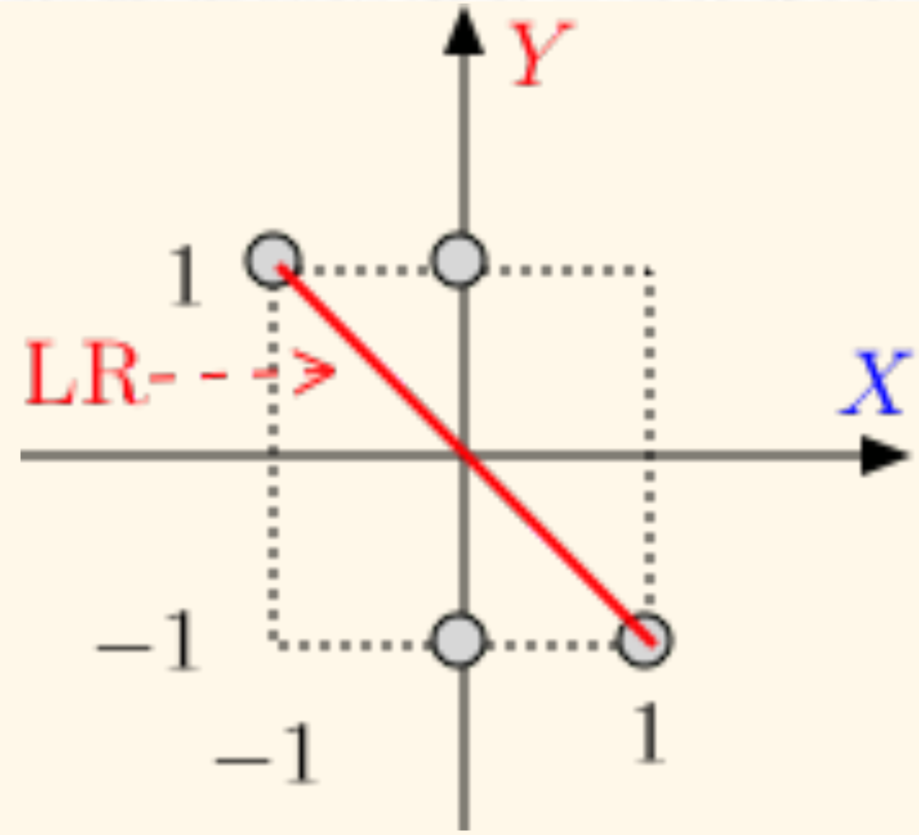
$\mathbb{E}(X^2) = \frac{1}{2} \quad \mathbb{E}(XY) = \frac{1}{2}$

$\mathrm{var}(X) = \frac{1}{2} \quad \mathrm{cov}(X,Y) = \frac{1}{2}.$

LR $Y = \mathbb{E}(Y) + \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}(X - \mathbb{E}(X)) = X$

Example 4

LR $\hat{Y} = -X$

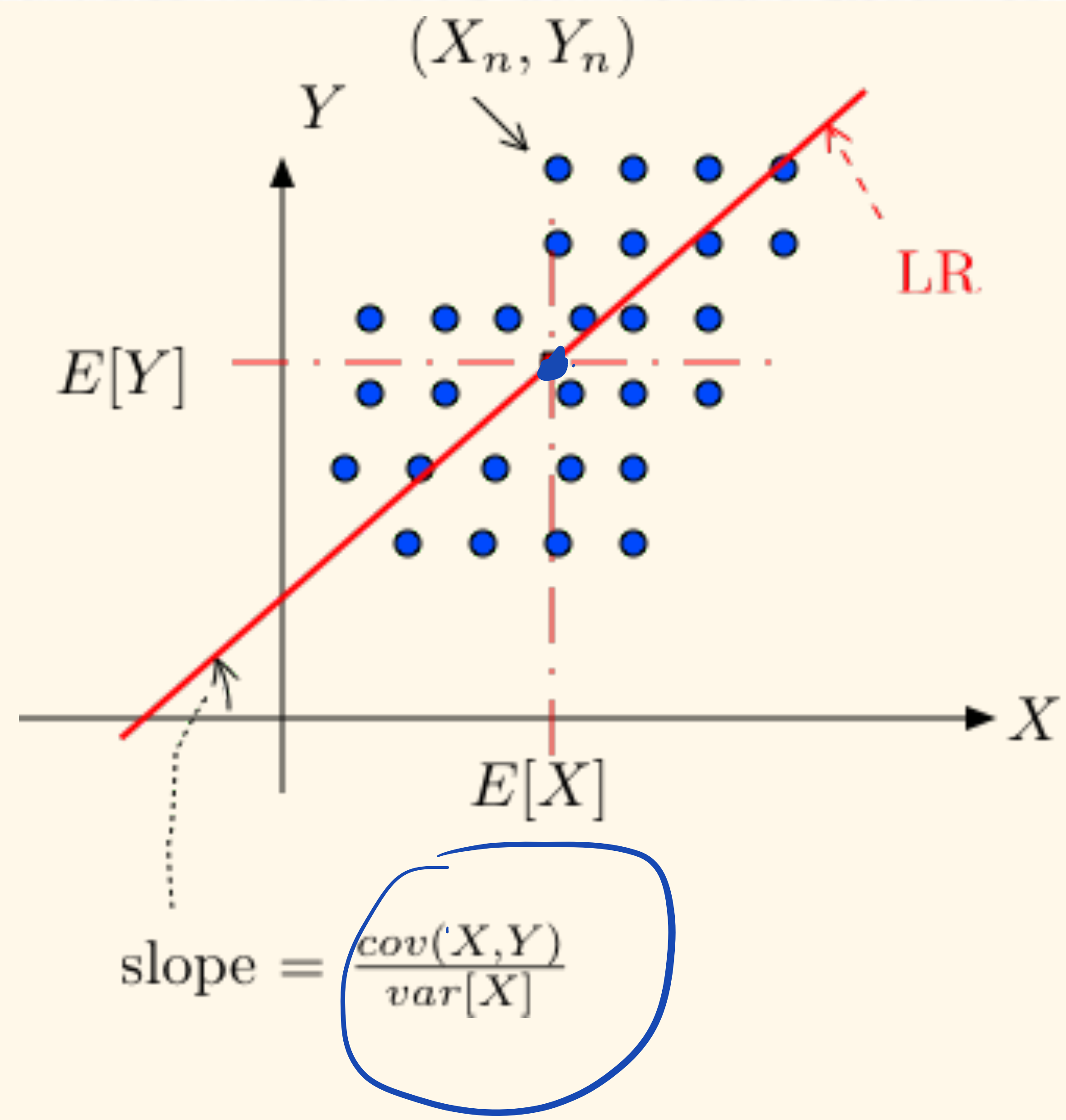


$\mathbb{E}(XY) = -\frac{1}{2}$

$\mathbb{E}(X) = 0$, $\mathbb{E}(Y) = 0$

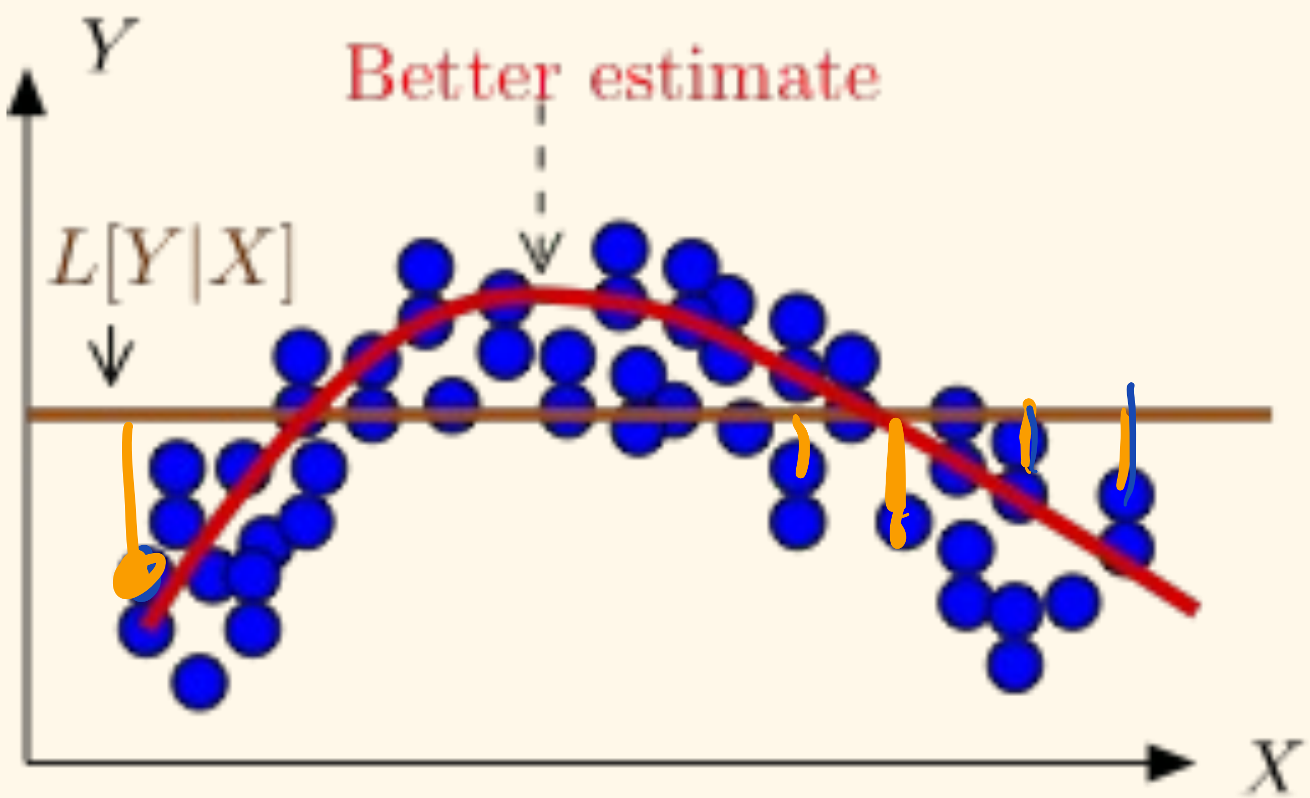
$\mathbb{E}(X^2) = \frac{1}{4}(1+0+0+1) = \frac{1}{2}$ $\quad Var(X) = \frac{1}{2}$

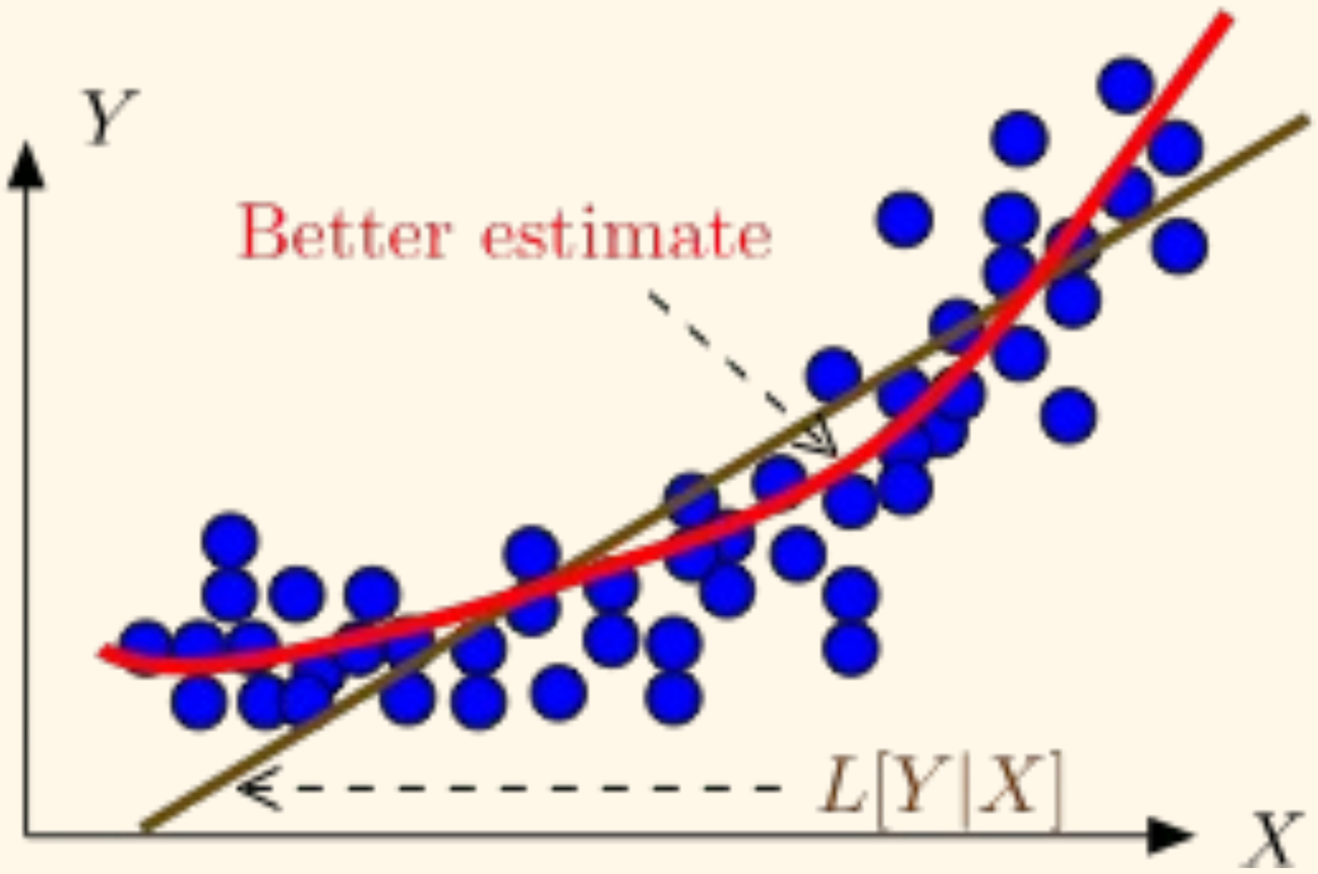
LR $\hat{Y} = -\frac{1}{2} / \frac{1}{2} X = -X$


(X_n, Y_n)
Y
LR
E[Y]
X
E[X]
slope = cov(X,Y) / var[X]

E(Y|X)

MMSE ...


Y
Better estimate
L[Y|X]
X


Y
Better estimate
L[Y|X]
X

Quick run through of probability

- Sample Space
- Random variable (discrete and continuous), function of r.v.
- distributions:
  - Uniform, Bernoulli, Binomial, Geometric, Poisson, Exponential, Normal, Piecewise constant …
  - Joint, marginal, conditional
- Bayes’ rule
- Expectation (conditional expectation)
- Variance, covariance, correlation, Independence
- Inequalities, WLLN, CLT
- Markov Chain
- MSE, LLSE formula

Final Tips

- Make use of cheat sheet
- Review definitions
- Easy points are easy to lose (silly mistake eats points..)
- Translate the question, statement or the quantity to find into math symbols
- Make valid assumption and write it down
- If you get stuck on one thing for too long, move on
  - z-score will help
- Be kind to your TA and ask for their tips
- Be honest and be proud