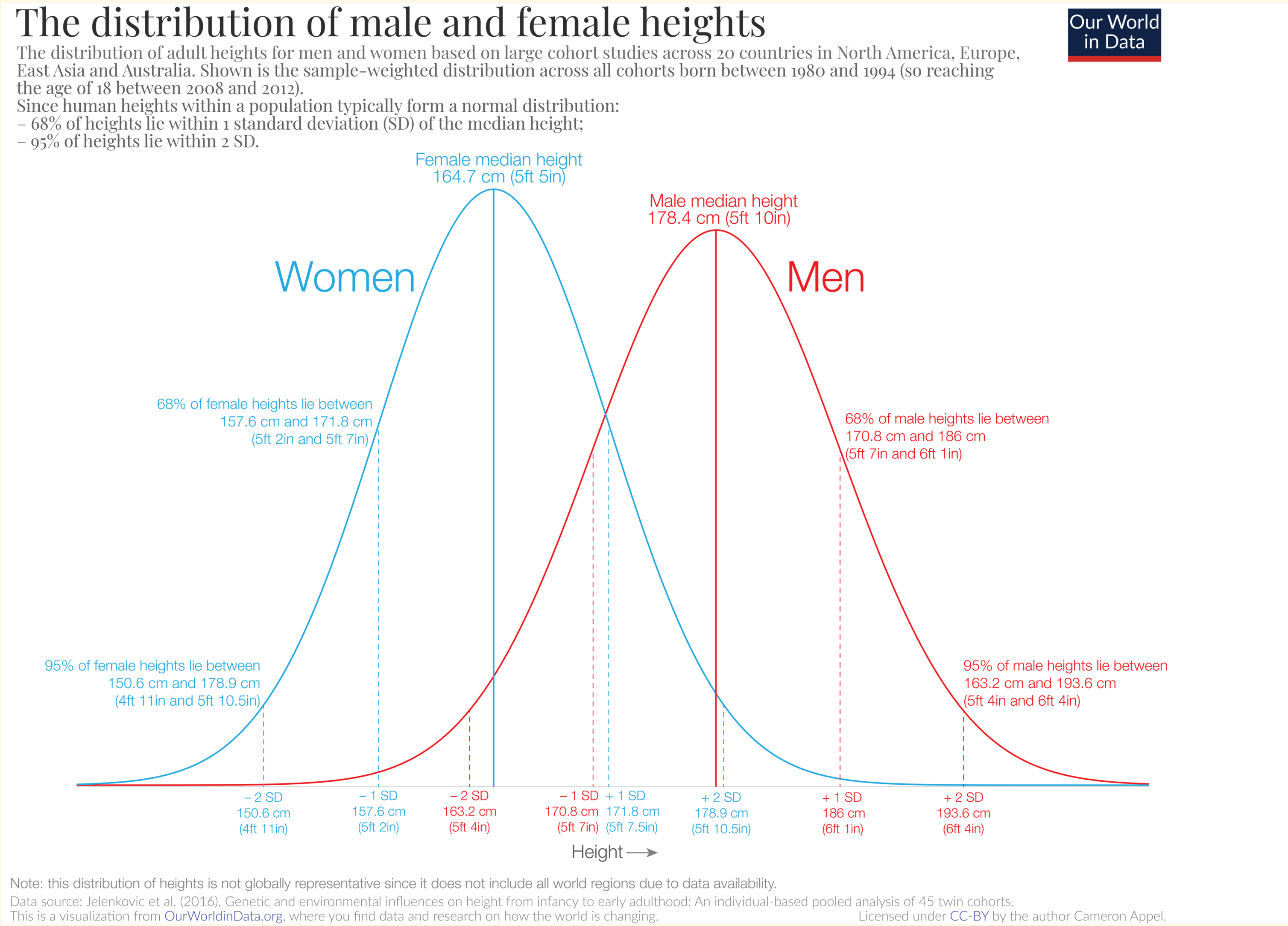
Estimation: LLSE

Aug 8, 2022

Uniform distribution

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

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.

Height of the person who sits next to you

- Now we have observation of this person’s weight.

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

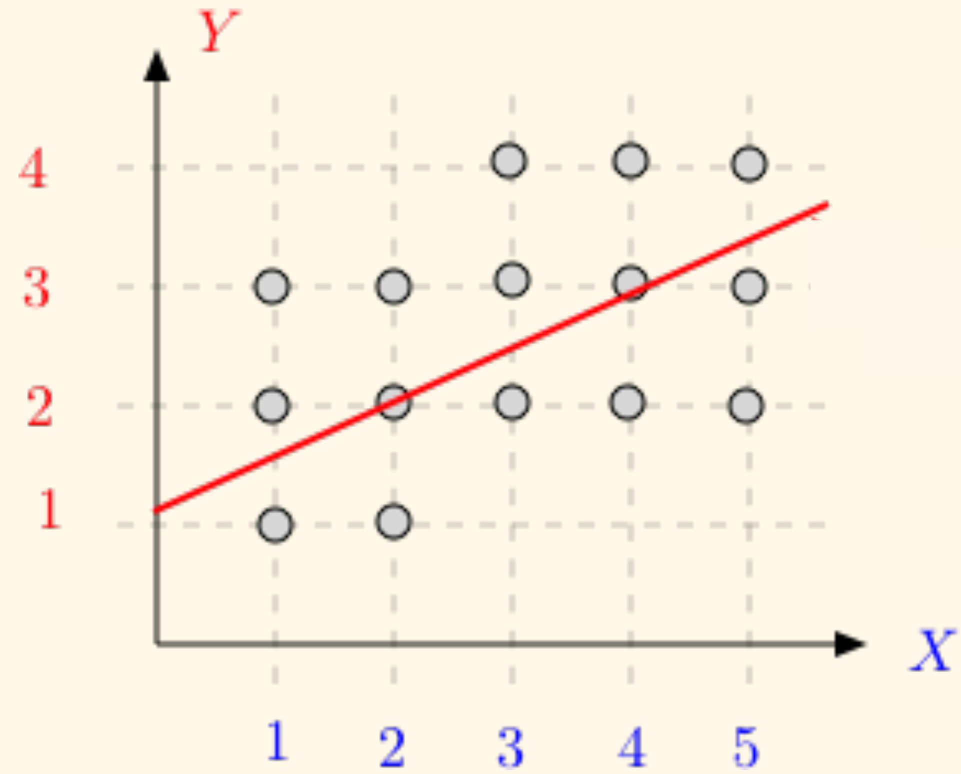
Linear least squared estimation

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

$$\mathbb{E}\left((Y - a - bX)^2\right)$$

Example 2.

Consider discrete joint distribution of X and Y



Linear regression

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

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

Linear regression of Y over X is

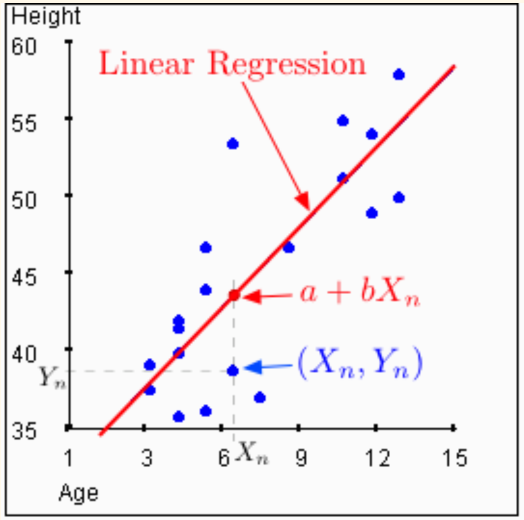
$$\hat{Y} = a + bX = \mathbb{E}(Y) + \frac{cov(X,Y)}{var(X)}(X - \mathbb{E}(X))$$

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

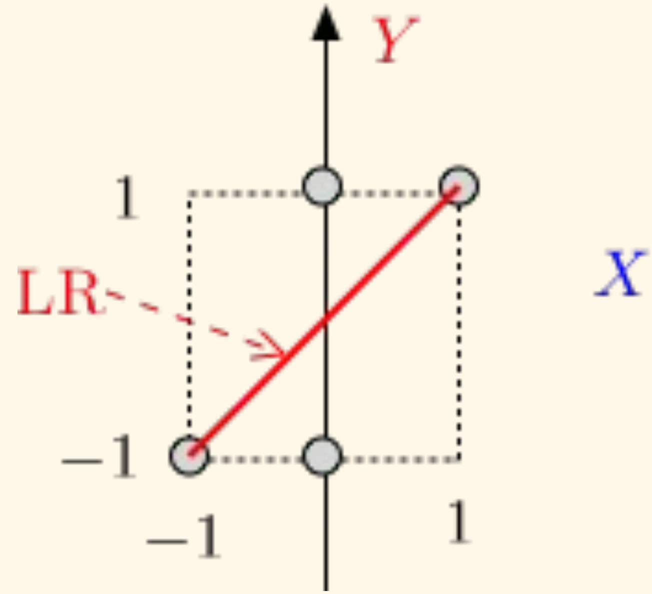
$var(X) = \frac{1}{K}\sum_{k=1}^{K}(X_k - \mathbb{E}(X))^2$, $cov(X,Y) = \frac{1}{K}\sum_{k=1}^{K} X_k Y_k - \mathbb{E}(X)\mathbb{E}(Y)$

Linear regression converge to LLSE

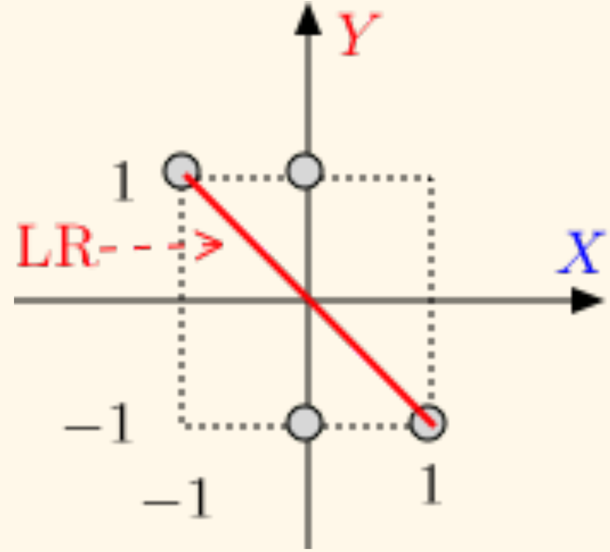
For X and Y, We observe K samples (X_1, Y_1) ... (X_K, Y_K). Assume that samples are i.i.d. As sample number increases, the linear regression approaches LLSE of X, Y.

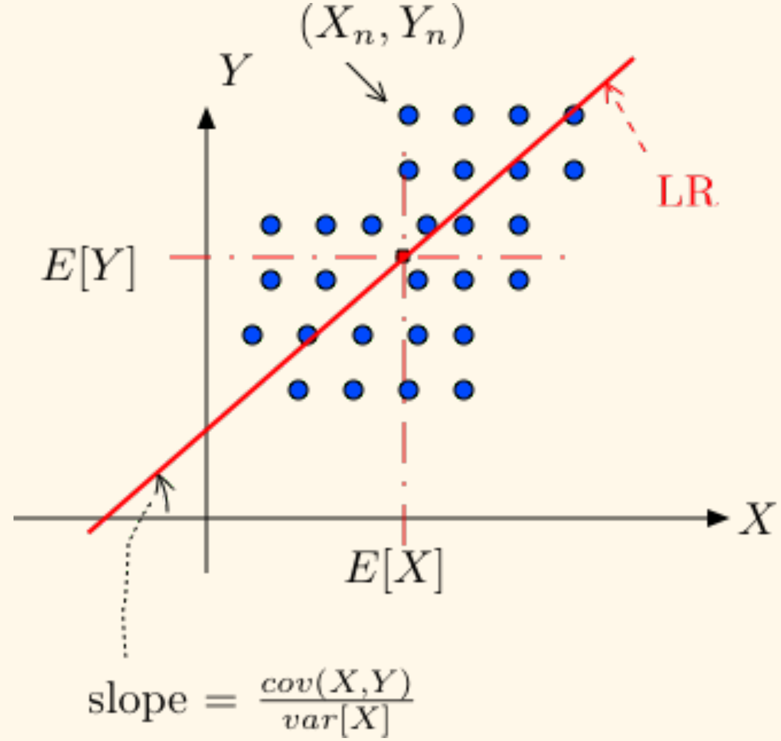


Example 3

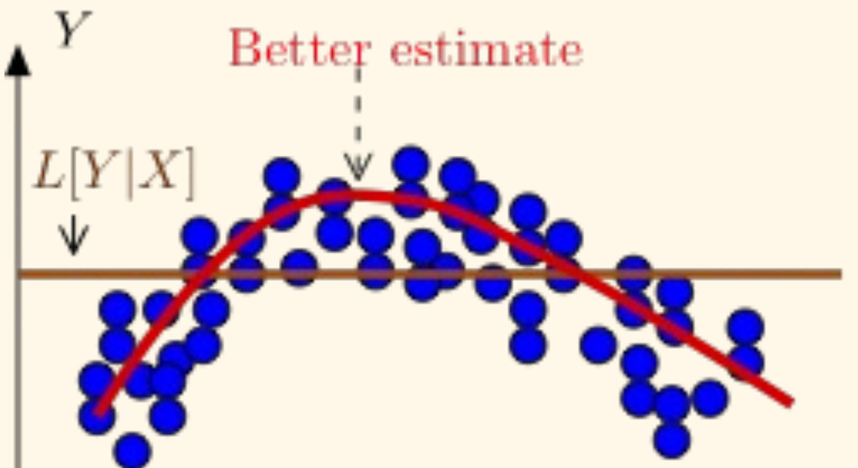


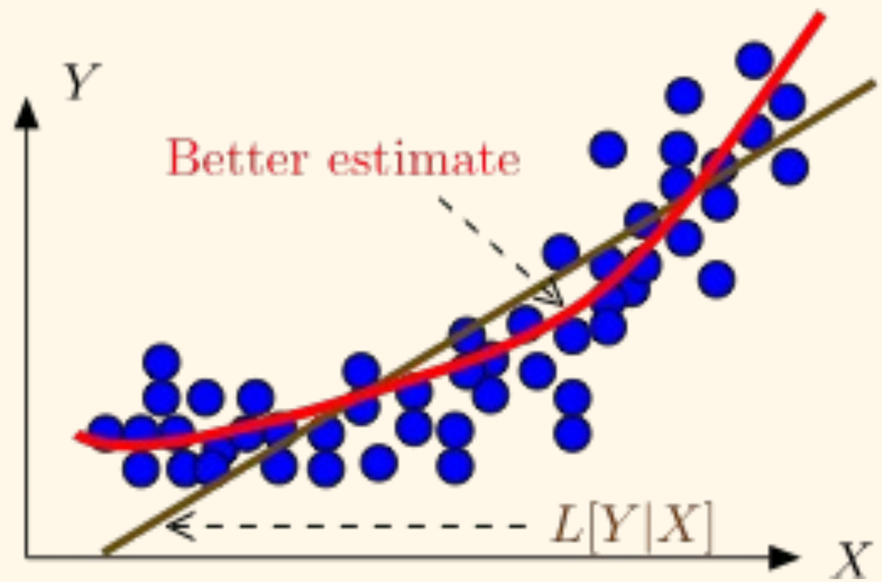
Example 4




(X_n, Y_n)
Y
LR
$E[Y]$
X
$E[X]$
slope $= \frac{cov(X,Y)}{var[X]}$

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