

Estimation: LLSE

Aug 8, 2022

Uniform distribution

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

Height of the person who sits next to you

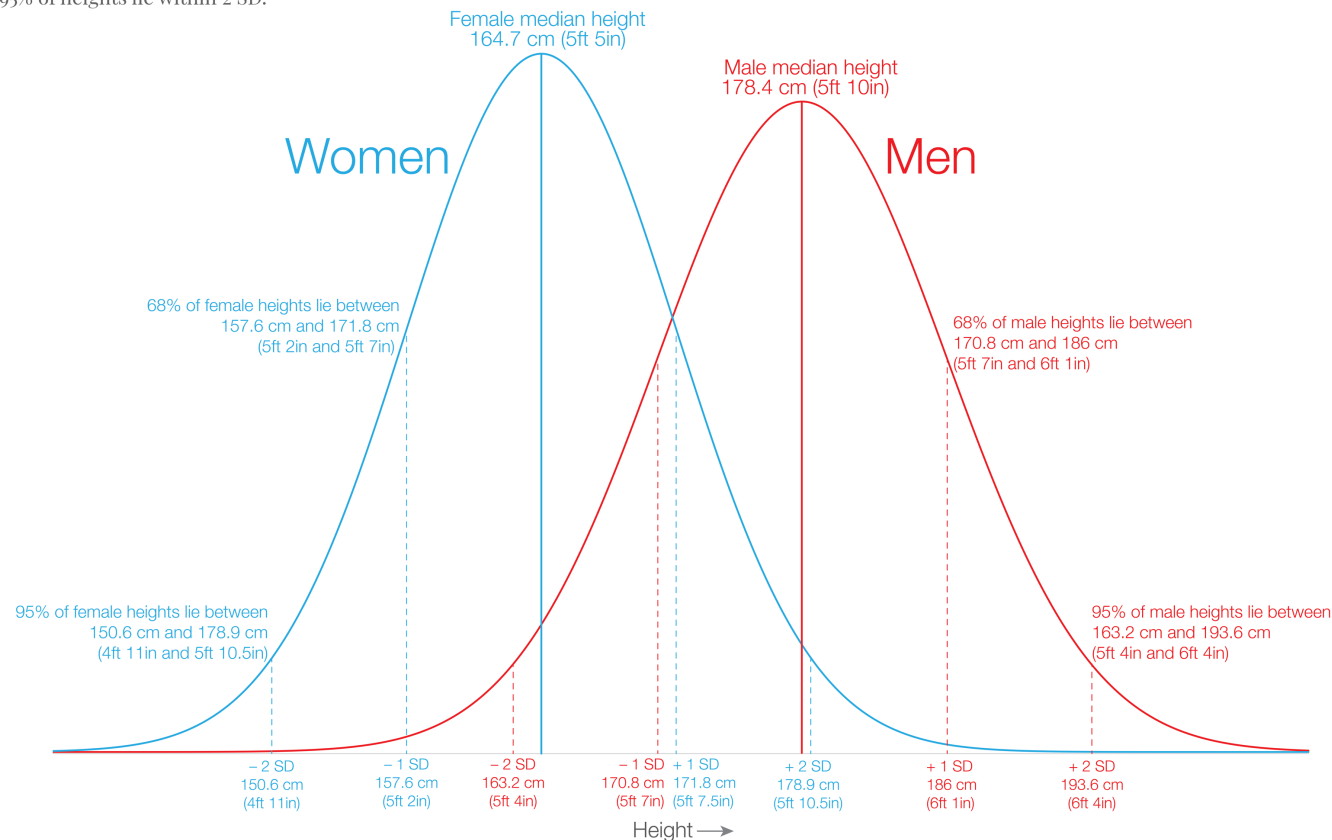
The distribution of male and female heights

Our World
in Data

The distribution of adult heights for men and women based on large cohort studies across 20 countries in North America, Europe, East Asia and Australia. Shown is the sample-weighted distribution across all cohorts born between 1980 and 1994 (so reaching the age of 18 between 2008 and 2012).

Since human heights within a population typically form a normal distribution:

- 68% of heights lie within 1 standard deviation (SD) of the median height;
- 95% of heights lie within 2 SD.



Note: this distribution of heights is not globally representative since it does not include all world regions due to data availability.

Data source: Jelenkovic et al. (2016). Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts.

This is a visualization from OurWorldinData.org, where you find data and research on how the world is changing.

Licensed under CC-BY by the author Cameron Appel.

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

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

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

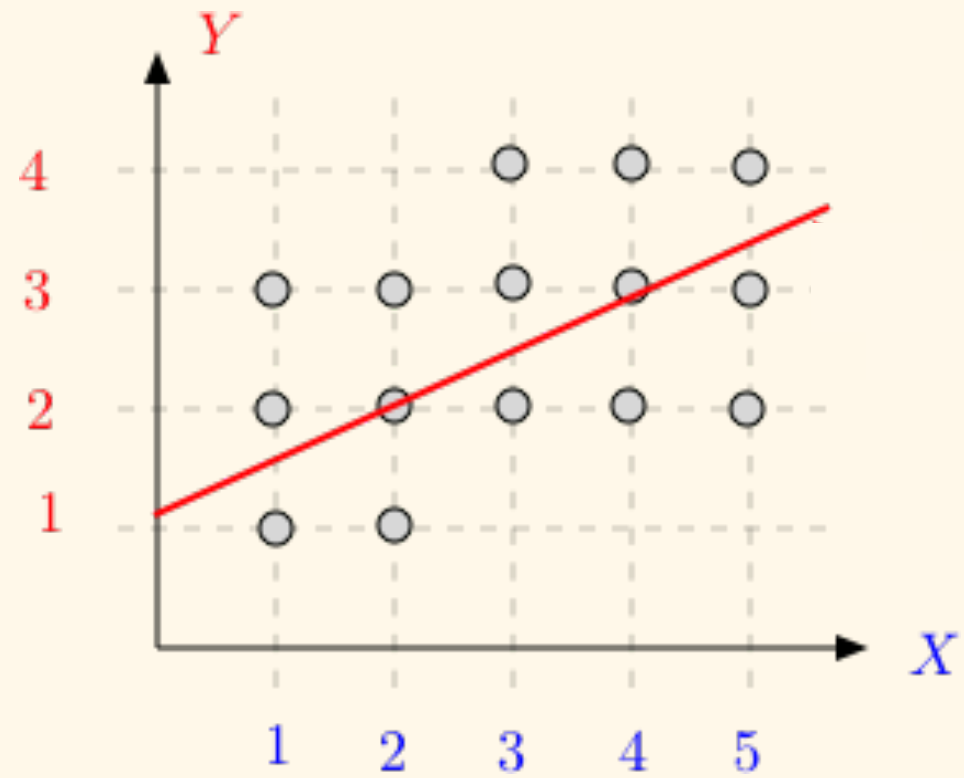
Linear least squared estimation

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

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

Example 2.

Consider discrete joint distribution of X and Y



Linear regression

For X and Y , We observe K samples $(X_1, Y_1) \dots (X_K, Y_K)$. $\hat{Y}_n = a + bX_n$ 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

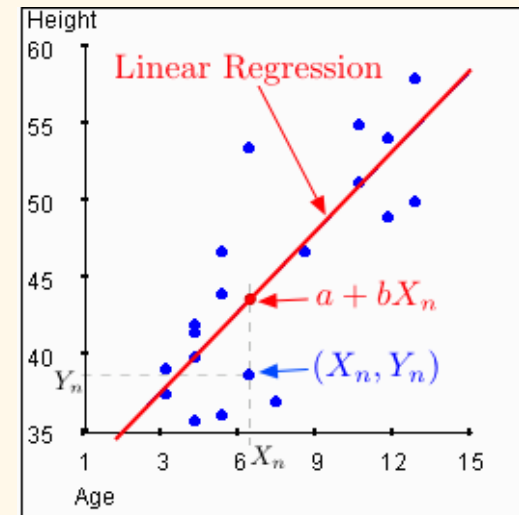
$$\hat{Y} = a + bX = \mathbb{E}(Y) + \frac{\text{cov}(X, Y)}{\text{var}(X)} (X - \mathbb{E}(X))$$

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

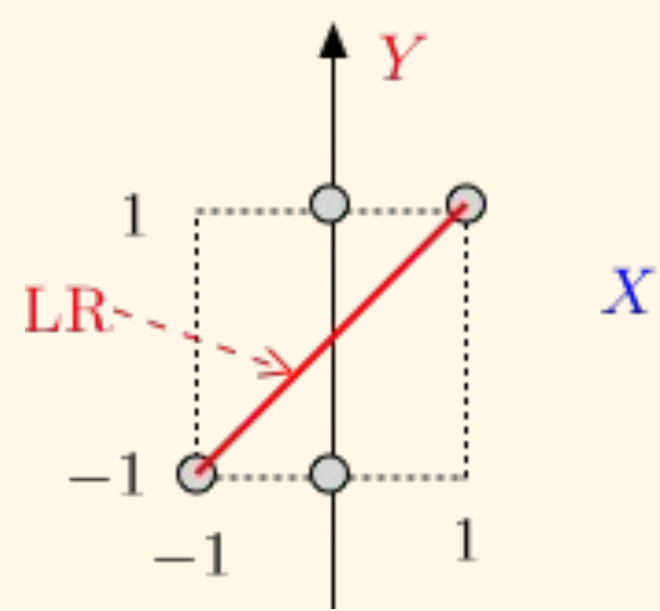
$\text{var}(X) = \frac{1}{K} \sum_{k=1}^K (X_k - \mathbb{E}(X))^2$, $\text{cov}(X, Y) = \frac{1}{K} \sum_{k=1}^K X_k Y_k - \mathbb{E}(X)\mathbb{E}(Y)$

Linear regression converge to LLSE

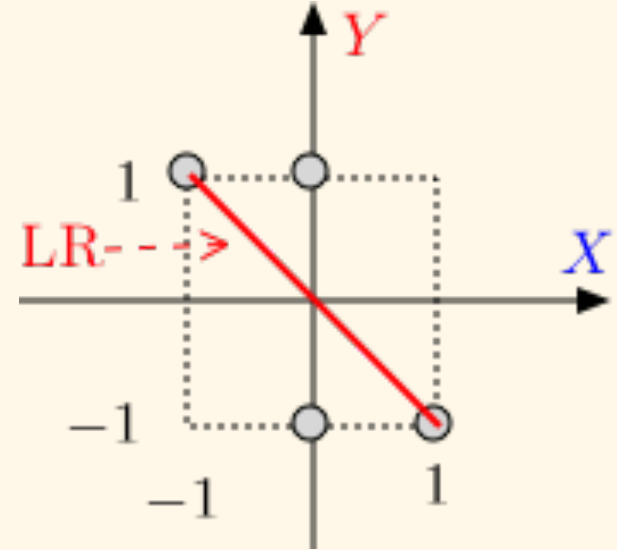
For X and Y , We observe K samples $(X_1, Y_1) \dots (X_K, Y_K)$.
Assume that samples are i.i.d. As sample number increases,
the linear regression approaches LLSE of X, Y .

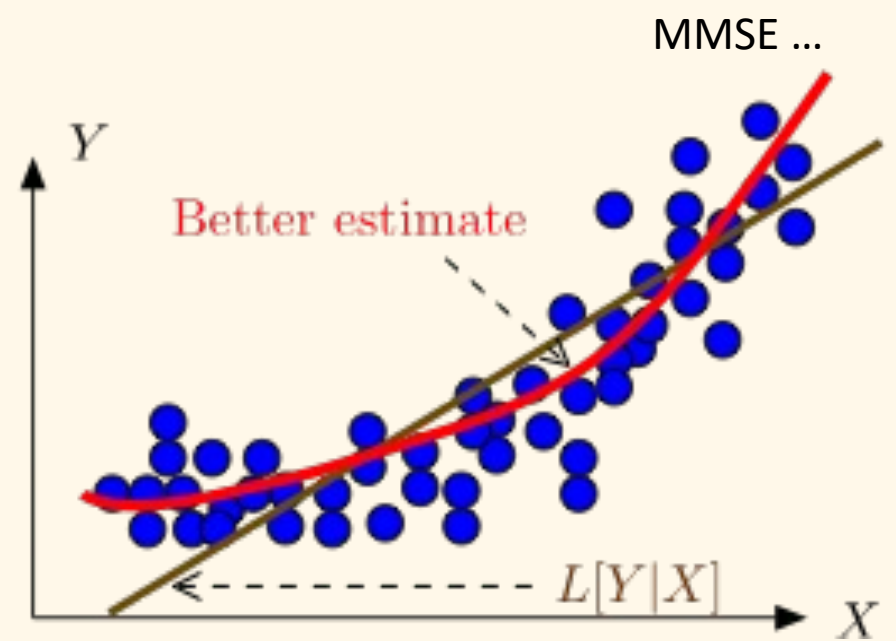
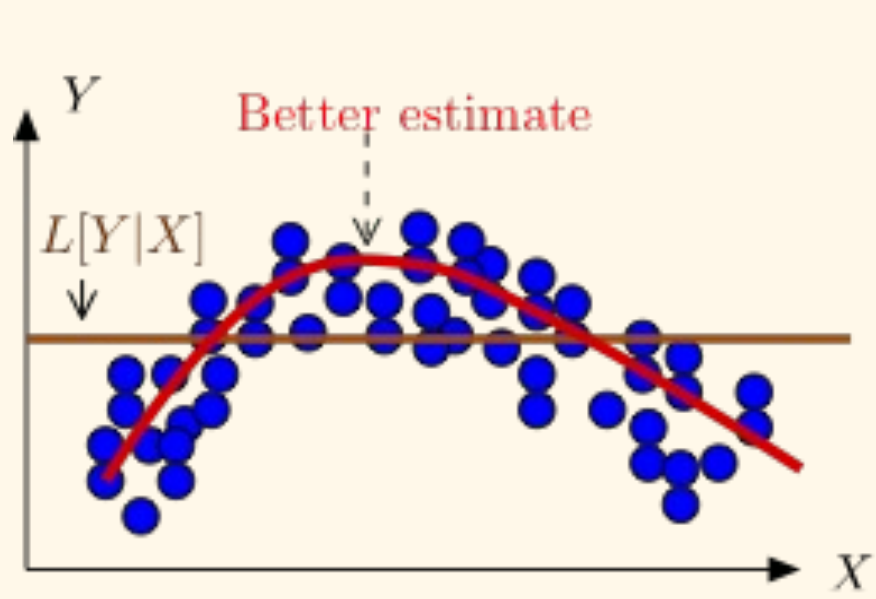
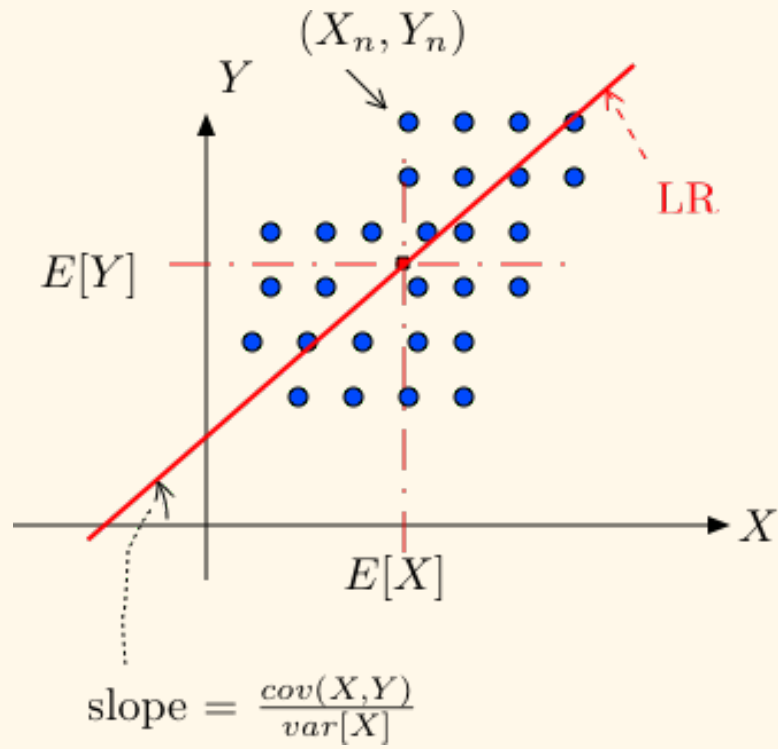


Example 3



Example 4





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