# Linear Estimation

- Recall the scalar estimation framework:
  - → There is an unobserved random variable $X$ and an observed random variable $Y$.
  - → An estimation rule $\hat{x}(Y)$ predicts the value of $X$ using only $Y$.
  - → The average quality of this prediction is measured by its mean-squared error $\mathbb{E}[(X-\hat{x}(Y))^2]$.
- We know that the optimal performance is attained by the minimum mean square error (MMSE) estimator, which corresponds to the conditional expectation $\hat{x}_{MMSE}(Y) = \mathbb{E}[X|Y]$.
  - → Unfortunately, it can be challenging to determine $\hat{x}_{MMSE}(Y)$, which is often a non-linear function of $Y$.
- What is the best possible linear estimator of the form $\hat{x}(Y) = aY + b$?

- The linear least squares error (LLSE) estimator $\hat{X}_{LLSE}(y)$ attains the smallest possible mean-squared error among all linear estimators:

$$\boxed{\begin{aligned}\hat{X}_{LLSE}(y) &= \mu_X + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\\ &= \mathbb{E}[X] + \frac{\mathrm{Cov}[X,Y]}{\mathrm{Var}[Y]}(y-\mathbb{E}[Y])\end{aligned}}$$

→ The mean-squared error (MSE) of the LLSE estimator is

$$\boxed{\begin{aligned}MSE_{LLSE} &= \sigma_X^2(1-\rho_{X,Y}^2)\\ &= \mathrm{Var}[X] - \frac{(\mathrm{Cov}[X,Y])^2}{\mathrm{Var}[Y]}\end{aligned}}$$

→ Note that, for jointly Gaussian X and Y, the MMSE and LLSE estimators are identical.

→ Determining the LLSE estimator is often simpler in practice, since we only need first- and second-order statistics.

- **Properties of the LLSE Estimator:**

→ The LLSE estimator is unbiased: $\mathbb{E}[\hat{X}_{LLSE}(Y)] = \mathbb{E}[X]$.

Why? $$\mathbb{E}[\hat{X}_{LLSE}(Y)] = \mathbb{E}\left[\mathbb{E}[X] + \frac{\text{Cov}[X,Y]}{\text{Var}[Y]}(Y - \mathbb{E}[Y])\right]$$
$$= \mathbb{E}[X] + \frac{\text{Cov}[X,Y]}{\text{Var}[Y]}(\mathbb{E}[Y] - \mathbb{E}[Y]) = \mathbb{E}[X]$$

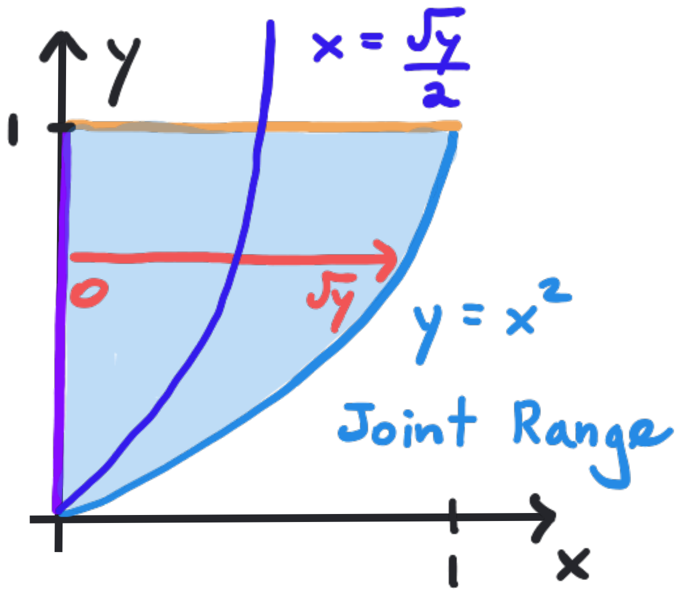
→ The error of the LLSE estimator is orthogonal to any linear function $aY + b$ of the observation:

$$\mathbb{E}\left[(X - \hat{X}_{LLSE}(Y))(aY + b)\right] = 0$$

See lecture notes for why this holds.

→ Another way to derive the LLSE estimator is to first establish that it must satisfy these two properties and then use them as a system of linear equations to solve for the LLSE coefficients.

- **Example:** $f_{X,Y}(x,y) = \begin{cases} \frac{3}{2} & 0 \le x \le \sqrt{y},\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$




→ What is the MMSE estimator?

$$\hat{x}_{MMSE}(y) = \mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\,dx$$

$$= \int_{0}^{\sqrt{y}} x \frac{1}{\sqrt{y}}\,dx = \left(\frac{1}{2}x^2\right)\Big|_0^{\sqrt{y}} \cdot \frac{1}{\sqrt{y}} = \frac{\sqrt{y}}{2}$$

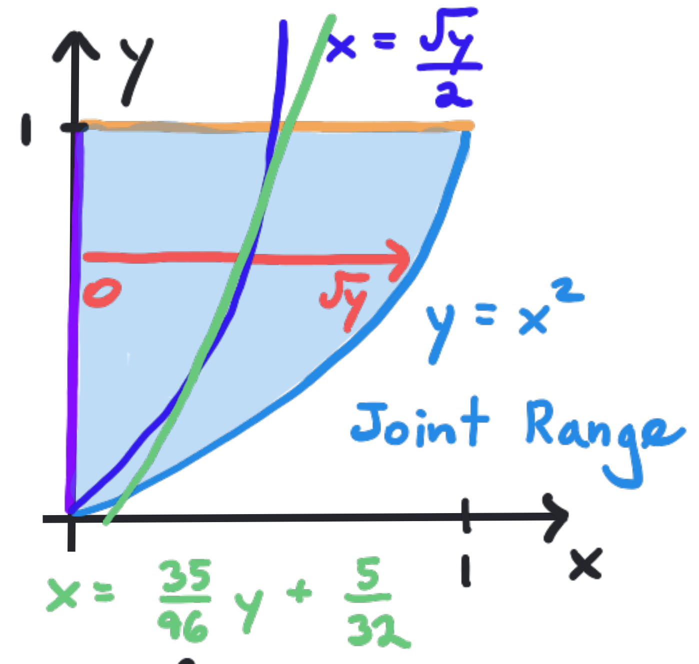
$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & (x,y) \in R_{X,Y} \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{\cancel{\frac{3}{2}}}{\cancel{\frac{3}{2}}\sqrt{y}} & 0 \le x \le \sqrt{y},\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{0}^{\sqrt{y}} \frac{3}{2}\,dx = \frac{3}{2}(x)\Big|_0^{\sqrt{y}} = \begin{cases} \frac{3}{2}\sqrt{y} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

→ What is its mean-squared error?

$$MSE_{MMSE} = \mathbb{E}\left[(X - \hat{x}_{MMSE}(Y))^2\right] = \int_0^1\int_0^{\sqrt{y}} \left(x - \frac{\sqrt{y}}{2}\right)^2 \frac{3}{2}\,dx\,dy \overset{\text{Computer}}{=} \frac{1}{20} = 0.05$$

• **Example:** $f_{X,Y}(x,y) = \begin{cases} \frac{3}{2} & \boxed{0 \le \boxed{x}} \le \sqrt{y}, \ 0 \le \boxed{y \le 1} \\ 0 & \text{otherwise} \end{cases}$




→ What is the LLSE estimator?

$$\hat{x}_{LLSE}(y) = \mathbb{E}[X] + \frac{Cov[X,Y]}{Var[Y]}(y - \mathbb{E}[Y])$$

$$\mathbb{E}[X] = \int_0^1\int_0^{\sqrt{y}} x \cdot \frac{3}{2}\,dx\,dy = \frac{3}{8}$$

$$\mathbb{E}[Y] = \int_0^1\int_0^{\sqrt{y}} y \cdot \frac{3}{2}\,dx\,dy = \frac{3}{5}$$

$$\mathbb{E}[Y^2] = \int_0^1\int_0^{\sqrt{y}} y^2 \cdot \frac{3}{2}\,dx\,dy = \frac{3}{7}$$

$$\mathbb{E}[XY] = \int_0^1\int_0^{\sqrt{y}} xy \cdot \frac{3}{2}\,dx\,dy = \frac{1}{4}$$

$$\mathbb{E}[X^2] = \int_0^1\int_0^{\sqrt{y}} x^2 \cdot \frac{3}{2}\,dx\,dy = \frac{1}{5}$$

$$Var[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$$
$$= \frac{3}{7} - \left(\frac{3}{5}\right)^2 = \frac{12}{175}$$

$$Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$
$$= \frac{1}{4} - \frac{3}{8} \cdot \frac{3}{5} = \frac{1}{40}$$

$$\hat{x}_{LLSE}(y) = \frac{3}{8} + \frac{1/40}{12/175}\left(y - \frac{3}{5}\right)$$
$$= \frac{35}{96}y + \frac{5}{32}$$

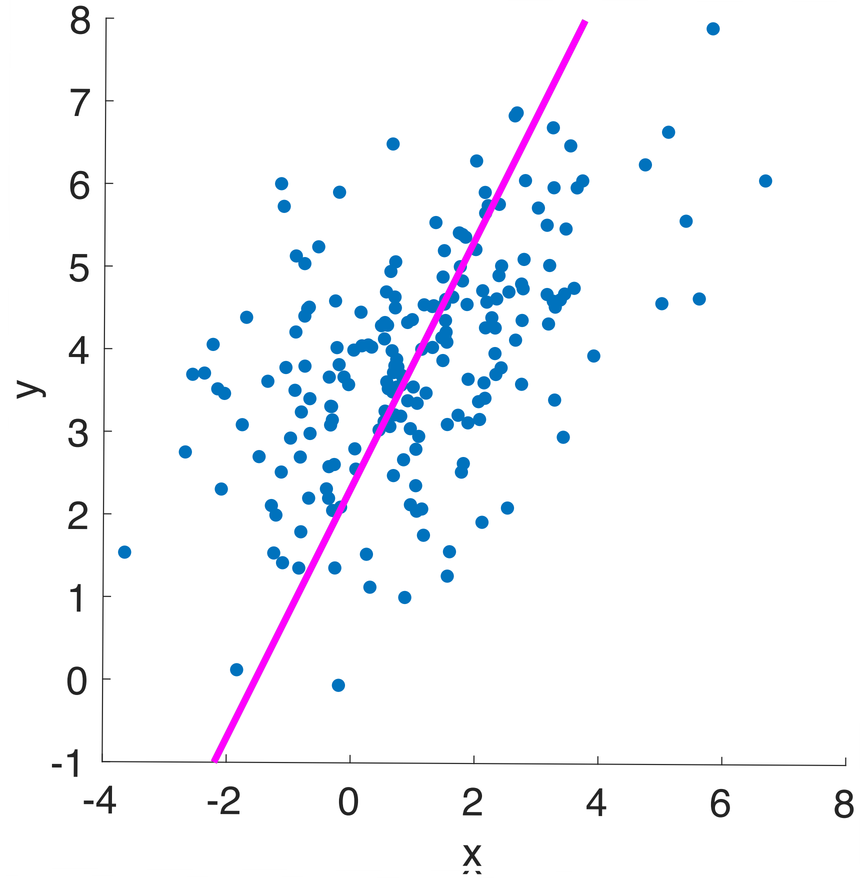
→ What is its mean-squared error?

$$MSE_{LLSE} = \mathbb{E}[(X - \hat{x}_{LLSE}(Y))^2] = Var[X] - \frac{(Cov[X,Y])^2}{Var[Y]} = \frac{19}{320} - \frac{(1/40)^2}{12/175} = \frac{193}{3840}$$

$$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320}$$

$$\approx 0.0503$$

- The LLSE estimator is frequently applied to real datasets where it is usually referred to as (simple) linear regression.




→ Dataset: $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$

→ Estimate means, variances, and covariance.

## Sample Means

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} X_i = 1.04 \qquad \hat{\mu}_Y = \frac{1}{n}\sum_{i=1}^{n} Y_i = 3.91$$

## Sample Variances

($\frac{1}{n-1}$ instead of $\frac{1}{n}$ makes unbiased)

$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_X)^2 = 2.96 \qquad \hat{\sigma}_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\mu}_Y)^2 = 1.89$$

## Sample Covariance

$$\widehat{\mathrm{Cov}}[X,Y] = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y) = 1.21$$

## Sample Correlation Coefficient

$$\hat{\rho}_{X,Y} = \frac{\widehat{\mathrm{Cov}}[X,Y]}{\hat{\sigma}_X \hat{\sigma}_Y} = 0.52$$

## Linear Regression Model

$$\hat{x}(y) = \hat{\mu}_X + \frac{\widehat{\mathrm{Cov}}[X,Y]}{\hat{\sigma}_Y^2}(y - \hat{\mu}_Y) = \hat{\mu}_X + \hat{\rho}_{X,Y}\frac{\hat{\sigma}_X}{\hat{\sigma}_Y}(y - \hat{\mu}_Y) = 0.66y - 1.54$$