# MMSE Estimation

- Probability is also an excellent foundation for making inferences from partial, noisy observations. This is known as estimation theory or statistical inference.

- Key Idea: Estimate the values of a set of unobserved random variables using the values of a set of observed random variables.

  → Ex: Finding the location of a target based on radar measurements.

  → Ex: Estimating the heart rate of a patient using electrical measurements.

  → Ex: Identifying the model parameters of an aerial drone from flight test data.

- We start with the scalar case to build intuition.

- **Scalar Estimation Framework:**

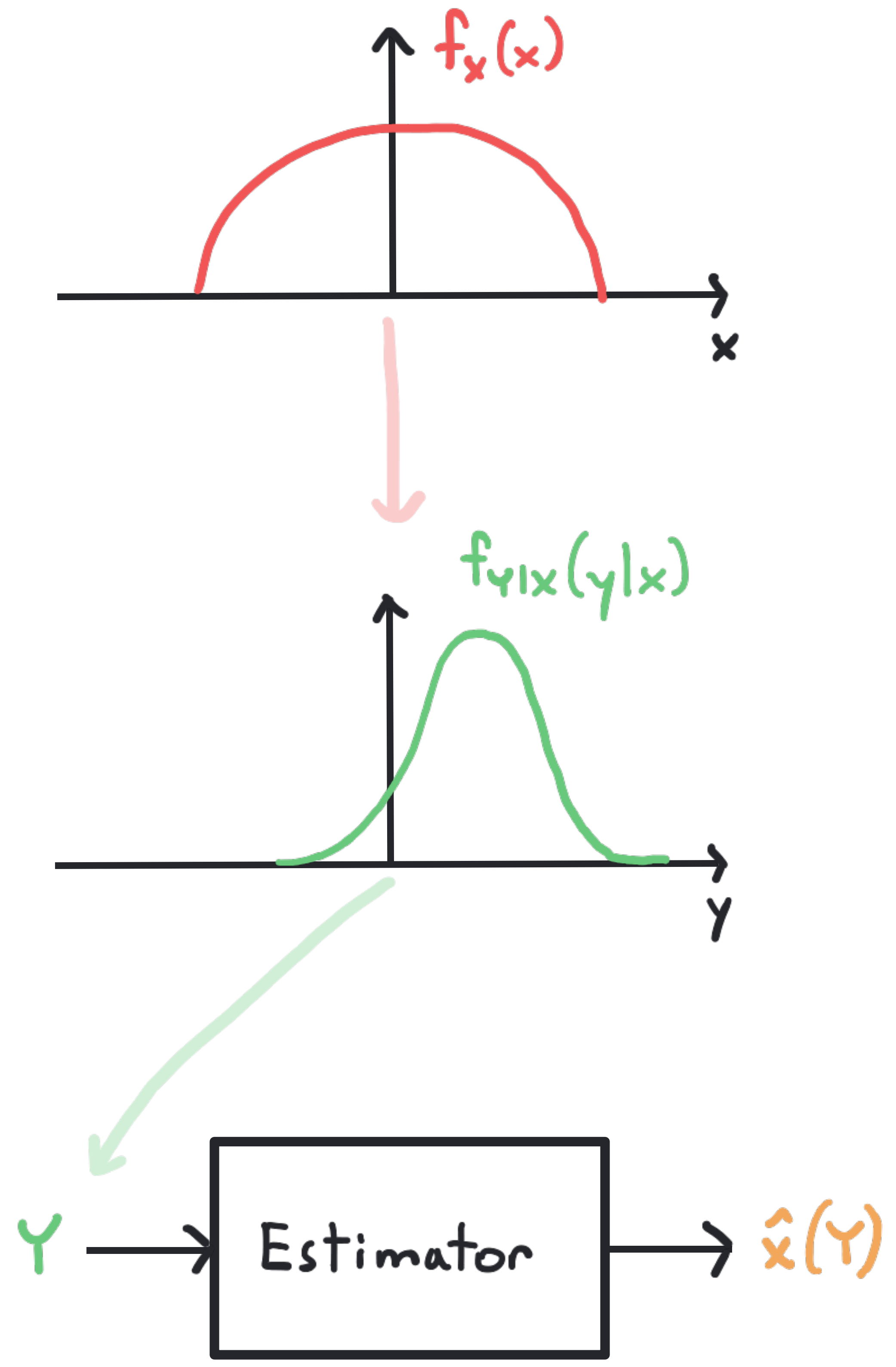
→ There is a prior distribution, which is the marginal distribution of the unobserved random variable $X$:

| Discrete Case | Continuous Case |
| --- | --- |
| $P_X(x)$ | $f_X(x)$ |

→ There is an observation model, which is the conditional distribution for the observed random variable $Y$:

| Discrete Case | Continuous Case |
| --- | --- |
| $P_{Y\|X}(y\|x)$ | $f_{Y\|X}(y\|x)$ |

→ There is an estimation rule $\hat{x}(Y)$, a function that outputs an estimate of the unobserved random variable.



- Recall that we used the probability of error as a measure of performance for detection.

- In most estimation problems, our estimate $\hat{x}(Y)$ will never be exactly equal to $X$. In these settings, $P_e = \mathbb{P}[\hat{x}(Y) \neq X] = 1$ for any choice of estimator. Therefore, we need a different measure of performance to compare estimators.

- In this class, we will focus exclusively on the mean-squared error MSE where

$$\boxed{\mathrm{MSE} = \mathbb{E}\left[(X - \hat{x}(Y))^2\right]}$$

$X - \hat{x}(Y)$ is called the error.

- We will also be interested in the bias of an estimator. Specifically, we say the estimator $\hat{x}(Y)$ is unbiased if the error $X - \hat{x}(Y)$ has zero mean, $\mathbb{E}[X - \hat{x}(Y)] = 0$.

- See your lecture notes for definitions and examples of the ML and MAP estimators, which are not optimal for MSE.

- The minimum mean square error (MMSE) estimator $\hat{x}_{MMSE}(y)$ attains the smallest possible MSE and is equal to the conditional expectation of $X$ given $Y=y$

$$\boxed{\hat{x}_{MMSE}(y) = \mathbb{E}[X \mid Y=y]}$$

→ Why? $\mathbb{E}[(X-\hat{x}(Y))^2] = \mathbb{E}[\mathbb{E}[(X-\hat{x}(Y))^2 \mid Y]]$ Law of Total Expectation

$\mathbb{E}[(X-\hat{x}(y))^2 \mid Y=y]$ Determine $\hat{x}(y)$ with the smallest error for each $y$.

$$= \mathbb{E}[X^2 - 2\hat{x}(y)X + \hat{x}^2(y) \mid Y=y]$$

Add and subtract this term to complete the square.

$$= \mathbb{E}[X^2 \mid Y=y] - 2\hat{x}(y)\mathbb{E}[X \mid Y=y] + \hat{x}^2(y)$$

$$= \mathbb{E}[X^2 \mid Y=y] - (\mathbb{E}[X \mid Y=y])^2 + (\mathbb{E}[X \mid Y=y])^2 - 2\hat{x}(y)\mathbb{E}[X \mid Y=y] + \hat{x}^2(y)$$

$$= \mathbb{E}[X^2 \mid Y=y] - (\mathbb{E}[X \mid Y=y])^2 + \boxed{(\mathbb{E}[X \mid Y=y] - \hat{x}(y))^2}$$

This is the only term that we can control.
Attains its minimum of 0 if and only if $\hat{x}(y) = \mathbb{E}[X \mid Y=y]$.

- <u>Properties of the MMSE Estimator</u>:

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

Why? $\mathbb{E}[\hat{x}_{MMSE}(Y)] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$

→ The error of the MMSE estimator is orthogonal to any function $g(Y)$ of the observation: $\mathbb{E}[(X - \hat{x}_{MMSE}(Y))g(Y)] = 0$.

Why? $\mathbb{E}[(X - \hat{x}_{MMSE}(Y))g(Y)]$

$= \mathbb{E}[\mathbb{E}[(X - \hat{x}_{MMSE}(Y))g(Y)|Y]]$ Law of Total Expectation

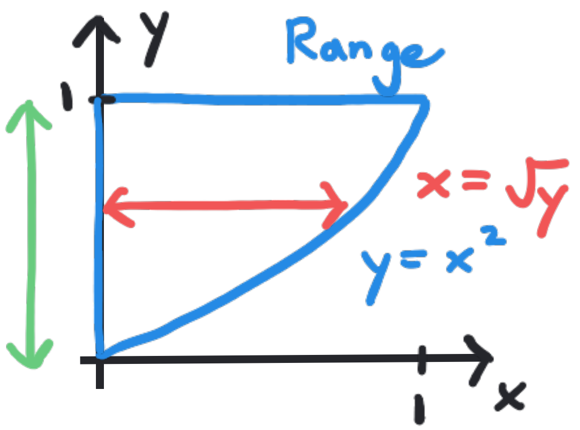
$= \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X|Y])g(Y)|Y]]$

$= \mathbb{E}[(\mathbb{E}[X|Y] - \mathbb{E}[X|Y])g(Y)]$ Linearity of Expectation

$= 0$

→ One useful consequence of orthogonality is that $\mathbb{E}[(X - \hat{x}_{MMSE}(Y))\hat{x}_{MMSE}(Y)] = 0$, which can be used to show that $\mathbb{E}[(X - \hat{x}_{MMSE}(Y))^2] = \mathbb{E}[X^2] - \mathbb{E}[\hat{x}^2_{MMSE}(Y)]$.

• **Example:** $f_{X,Y}(x,y) = \begin{cases} \frac{12}{11}(x+1) & 0 \le x \le \sqrt{y},\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$




→ Determine the MMSE estimator.

$$\hat{x}_{MMSE}(y) = \mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\,dx = \int_{0}^{\sqrt{y}} x\, \frac{2(x+1)}{y + 2\sqrt{y}}\,dx \overset{\text{computer}}{=} \frac{2y + 3\sqrt{y}}{3\sqrt{y} + 6}$$

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & (x,y) \in R_{X,Y} \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \boxed{\dfrac{\frac{12}{11}(x+1)}{\frac{6}{11}(y + 2\sqrt{y})}} & 0 \le x \le \sqrt{y},\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{0}^{\sqrt{y}} \frac{12}{11}(x+1)\,dx \overset{\text{Computer}}{=} \begin{cases} \boxed{\frac{6}{11}(y + 2\sqrt{y})} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

→ Determine the resulting mean-squared error (MSE).

$$MSE_{MMSE} = \mathbb{E}[(X - \hat{x}_{MMSE}(Y))^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \hat{x}_{MMSE}(y))^2 f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_{0}^{1}\int_{0}^{\sqrt{y}} \left(x - \frac{2y + 3\sqrt{y}}{3\sqrt{y} + 6}\right)^2 \frac{12}{11}(x+1)\,dx\,dy$$

Computer → $= \frac{46}{55} - \frac{64}{33}\ln\left(\frac{3}{2}\right) \approx 0.05$

- **Estimation for Jointly Gaussian Random Variables:**

→ If $X$ and $Y$ are jointly Gaussian, then the MMSE estimator is a linear function of $Y$:

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

→ The mean-squared error in this case is

$$\boxed{\begin{aligned} MSE_{MMSE} &= \mathbb{E}\left[(X - \hat{x}_{MMSE}(Y))^2\right] = (1 - \rho_{X,Y}^2)\,\sigma_X^2 \\ &= \mathrm{Var}[X] - \frac{(\mathrm{Cov}[X,Y])^2}{\mathrm{Var}[Y]} \end{aligned}}$$