7 Estimation

- Probability is also an excellent foundation for making inferences from noisy observations.
- Key Idea: Estimate the values of a set of *unobserved* random variables using the values of a set of *observed* random variables.

7.1 Scalar Estimation

- There is an *unobserved* random variable X, generated by a **prior distribution:**

$$\text{Discrete: } P_X(x) \qquad\qquad \text{Continuous: } f_X(x)$$

- There is an *observed* random variable Y, generated by an **observation model:**

$$\text{Discrete: } P_{Y|X}(y|x) \qquad\qquad \text{Continuous: } f_{Y|X}(y|x)$$

- We need to select an **estimation rule** (or **estimator**) $\hat{x}(y)$, which takes the value of the observed random variable $Y = y$ as an input and outputs an estimate of the unobserved random variable X.
- Since it is unlikely the estimate $\hat{x}(Y)$ will be exactly equal to X, it does not make sense to analyze the performance of our estimators with the probability of error. Instead, we need a criterion for measuring the error between the true X and our estimate $\hat{x}(Y)$. There are many ways to do this.
- Here, we focus on the **mean-squared error (MSE),** $\mathsf{MSE} = \mathbb{E}\big[(X - \hat{x}(Y))^2\big]$,

$$\text{Discrete: } \mathsf{MSE} = \sum_{x \in R_X} \sum_{y \in R_Y} \big(x - \hat{x}(y)\big)^2 P_{X,Y}(x,y) \qquad \text{Continuous: } \mathsf{MSE} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(x - \hat{x}(y)\big)^2 f_{X,Y}(x,y)\, dx\, dy$$

7.2 Minimum Mean-Squared Error (MMSE) Estimator

- The **minimum mean-squared error (MMSE) estimator** $\hat{x}_{\text{MMSE}}(y)$ is the estimator that attains the smallest mean-squared error across all possible estimators.
- It turns out that that the MMSE estimator is equivalent to the conditional expected value of X given $Y = y$,

$$\hat{x}_{\text{MMSE}}(y) = \mathbb{E}[X|Y = y].$$

- The MMSE estimator satisfies the following properties:
  - **Unbiased:** Its expected value is the same as that of the desired random variable,

  $$\mathbb{E}\big[\hat{x}_{\text{MMSE}}(Y)\big] = \mathbb{E}[X].$$

  - **Orthogonality Principle:** Its error is orthogonal to any function of the observed random variable,

  $$\mathbb{E}\Big[\big(X - \hat{x}_{\text{MMSE}}(Y)\big)g(Y)\Big] = 0 \qquad \text{for any function } g(y).$$

  An important special case is that the MMSE estimator is orthogonal to its own error,

  $$\mathbb{E}\Big[\big(X - \hat{x}_{\text{MMSE}}(Y)\big)\hat{x}_{\text{MMSE}}(Y)\Big] = 0.$$

7.3 Linear Least-Squares Error (LLSE) Estimator

- In some scenarios, it can be tricky to calculate the MMSE estimator. In these cases, we sometimes turn to the **linear least-squares error (LLSE) estimator**, which attains the smallest mean-squared error across all possible linear estimators. Here are two equivalent formulas for the LLSE estimator:

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

- The mean-squared error of the LLSE estimator is

$$\mathsf{MSE}_{\text{LLSE}} = \mathbb{E}\Big[\big(X - \hat{x}_{\text{LLSE}}(Y)\big)^2\Big] = \mathsf{Var}[X] - \frac{\big(\mathsf{Cov}[X,Y]\big)^2}{\mathsf{Var}[Y]} = \mathsf{Var}[X]\big(1 - \rho_{X,Y}^2\big)$$

- The LLSE estimator satisfies the following properties:
  - **Unbiased:** Its expected value is the same as that of the desired random variable,
    $$\mathbb{E}\big[\hat{x}_{\text{LLSE}}(Y)\big] = \mathbb{E}[X].$$
  - **Orthogonality Principle:** Its error is orthogonal to any linear function of the observed random variable,
    $$\mathbb{E}\big[\big(X - \hat{x}_{\text{LLSE}}(Y)\big)(aY+b)\big] = 0 \qquad \text{for any } a, b.$$
    An important special case is that the LLSE estimator is orthogonal to its own error,
    $$\mathbb{E}\big[\big(X - \hat{x}_{\text{LLSE}}(Y)\big)\hat{x}_{\text{LLSE}}(Y)\big] = 0.$$
  - For the special case of jointly Gaussian X and Y, the LLSE estimator is also the MMSE estimator.

7.4 Vector Estimation

- There is an *unobserved* random vector $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, generated by a **prior distribution:**

  Discrete: $P_{\underline{X}}(\underline{x})$ Continuous: $f_{\underline{X}}(\underline{x})$

- There is an *observed* random vector $\underline{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix}$, generated by an **observation model:**

  Discrete: $P_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$ Continuous: $f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$

- We need to select an **estimation rule** (or **estimator**) $\underline{\hat{x}}(\underline{y}) = \begin{bmatrix} \hat{x}_1(\underline{y}) \\ \vdots \\ \hat{x}_n(\underline{y}) \end{bmatrix}$.

- We measure the performance of our estimator by its **mean-squared error (MSE),**

$$\mathsf{MSE} = \sum_{i=1}^{n} \mathbb{E}\big[\big(X_i - \hat{x}_i(\underline{Y})\big)^2\big] = \mathbb{E}\big[\big(\underline{X} - \underline{\hat{x}}(\underline{Y})\big)^{\mathsf{T}}\big(\underline{X} - \underline{\hat{x}}(\underline{Y})\big)\big] \ .$$

7.5 Vector Minimum Mean-Squared Error (MMSE) Estimator

- The **(vector) minimum mean-squared error (MMSE) estimator** $\underline{\hat{x}}_{\text{MMSE}}(\underline{y})$ is the estimator that attains the smallest mean-squared error across all possible estimators,

$$\underline{\hat{x}}_{\text{MMSE}}(\underline{y}) = \mathbb{E}\big[\underline{X}|\underline{Y}=\underline{y}\big] = \begin{bmatrix} \mathbb{E}\big[X_1|\underline{Y}=\underline{y}\big] \\ \vdots \\ \mathbb{E}\big[X_n|\underline{Y}=\underline{y}\big] \end{bmatrix}$$

$$\mathbb{E}\big[X_i|\underline{Y}=\underline{y}\big] = \begin{cases} \displaystyle\sum_{x_i \in R_{X_i}} x_i\, P_{X_i|Y_1,\dots,Y_m}(x_i|y_1,\dots,y_m) & X_i \text{ discrete} \\ \displaystyle\int_{-\infty}^{\infty} x_i\, f_{X_i|Y_1,\dots,Y_m}(x_i|y_1,\dots,y_m)\,dx_i & X_i \text{ continuous} \end{cases}$$

- However, this formula can be difficult to evaluate, both analytically and empirically.

7.6 Vector Linear Least-Squares Error (LLSE) Estimator

- The **(vector) linear least-squares error (LLSE) estimator** is the estimator that attains the smallest mean-squared error across all possible linear estimators,

$$\underline{\hat{x}}_{\text{LLSE}}(\underline{y}) = \mathbb{E}[\underline{X}] + \mathbf{\Sigma}_{\underline{X},\underline{Y}}\mathbf{\Sigma}_{\underline{Y}}^{-1}\big(\underline{y} - \mathbb{E}[\underline{Y}]\big)\ ,$$

  which involves the

  - Covariance Matrix of $\underline{Y}$: $\mathbf{\Sigma}_{\underline{Y}} = \mathbb{E}\big[(\underline{Y}-\mathbb{E}[\underline{Y}])(\underline{Y}-\mathbb{E}[\underline{Y}])^{\mathsf{T}}\big] = \begin{bmatrix} \mathsf{Cov}[Y_1,Y_1] & \cdots & \mathsf{Cov}[Y_1,Y_m] \\ \vdots & \ddots & \vdots \\ \mathsf{Cov}[Y_m,Y_1] & \cdots & \mathsf{Cov}[Y_m,Y_m] \end{bmatrix}$
  - Cross-Covariance Matrix: $\mathbf{\Sigma}_{\underline{X},\underline{Y}} = \mathbb{E}\big[(\underline{X}-\mathbb{E}[\underline{X}])(\underline{Y}-\mathbb{E}[\underline{Y}])^{\mathsf{T}}\big] = \begin{bmatrix} \mathsf{Cov}[X_1,Y_1] & \cdots & \mathsf{Cov}[X_1,Y_m] \\ \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_n,Y_1] & \cdots & \mathsf{Cov}[X_n,Y_m] \end{bmatrix}$

- The mean-squared error of the LLSE estimator is

$$\mathsf{MSE}_{\text{LLSE}} = \mathsf{Tr}\big(\mathbf{\Sigma}_{\underline{X}} - \mathbf{\Sigma}_{\underline{X},\underline{Y}}\mathbf{\Sigma}_{\underline{Y}}^{-1}\mathbf{\Sigma}_{\underline{X},\underline{Y}}^{\mathsf{T}}\big)$$

  where $\mathbf{\Sigma}_{\underline{X}}$ is the covariance matrix of $\underline{X}$ and Tr is the trace operator, which sums up the diagonal elements of a matrix.

- The LLSE estimator satisfies the following properties:
  - **Unbiased:** Its expected value is the same as that of the desired random variable,

$$\mathbb{E}\big[\underline{\hat{x}}_{\text{LLSE}}(\underline{Y})\big] = \mathbb{E}[\underline{X}]$$

  - **Orthogonality Principle:** Its error is orthogonal to any linear function of the observed random variable,

$$\mathbb{E}\big[\big(\underline{X} - \underline{\hat{x}}_{\text{LLSE}}(\underline{Y})\big)(\mathbf{A}\underline{Y} + \underline{b})^{\mathsf{T}}\big] = \mathbf{0}$$

    for any linear function $\mathbf{A}\underline{y} + \underline{b}$. An important special case is that the LLSE estimator is orthogonal to its own error,

$$\mathbb{E}\big[\big(\underline{X} - \underline{\hat{x}}_{\text{LLSE}}(\underline{Y})\big)\underline{\hat{x}}_{\text{LLSE}}(\underline{Y})\big] = 0.$$

- For the special case where $\begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix}$ is a Gaussian vector, the vector LLSE estimator is also the vector MMSE estimator.