- **Vector Estimation Framework:**

→ In general, there may be one or more observed random variables $Y_1, \ldots, Y_m$ that we will use to estimate the values of one or more unobserved random variables $X_1, \ldots, X_n$.

→ Convenient to organize these into vectors: $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ $\underline{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix}$

(possibly of different lengths)

| | Discrete Case | Continuous Case |
| --- | --- | --- |
| → Prior Distribution: | $P_{\underline{X}}(\underline{x})$ | $f_{\underline{X}}(\underline{x})$ |
| → Observation Model: | $P_{\underline{Y}\mid\underline{X}}(\underline{y}\mid\underline{x})$ | $f_{\underline{Y}\mid\underline{X}}(\underline{y}\mid\underline{x})$ |

→ The estimation rule $\hat{\underline{x}}(\underline{Y})$ outputs a vector $\hat{\underline{x}}(\underline{Y}) = \begin{bmatrix} \hat{x}_1(\underline{Y}) \\ \vdots \\ \hat{x}_n(\underline{Y}) \end{bmatrix}$.

→ The mean-squared error MSE is

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

- The vector minimum mean-squared error (MMSE) estimator can be expressed using a compact formula,

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

where
$$\mathbb{E}[X_i \mid \underline{Y} = \underline{y}] = \begin{cases} \sum_{x_i \in R_{X_i}} x_i \, P_{X_i \mid Y_1, \ldots, Y_m}(x_i \mid y_1, \ldots, y_m) & X_i \text{ is Discrete} \\ \int_{-\infty}^{\infty} x_i \, f_{X_i \mid Y_1, \ldots, Y_m}(x_i \mid y_1, \ldots, y_m) \, dx_i & X_i \text{ is Continuous} \end{cases}$$

- Unfortunately, evaluating this formula for a specific distribution can be very difficult.
- Similarly, it can be hard to determine the vector MMSE estimator empirically from a dataset when we do not know the distribution.

- We can avoid these challenging computations by restricting ourselves to linear estimators of the form $\hat{\underline{x}}(\underline{y}) = \mathbf{A}\underline{y} + \underline{b}$ for some $n \times m$ matrix $\mathbf{A}$ and length-$n$ column vector $\underline{b}$.

- The vector linear least squares error (LLSE) estimator $\hat{\underline{x}}_{LLSE}(\underline{y})$ is

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

→ Recall the covariance matrix of $\underline{Y}$ is

$$\Sigma_{\underline{Y}} = \mathbb{E}[(\underline{Y} - \mathbb{E}[\underline{Y}])(\underline{Y} - \mathbb{E}[\underline{Y}])^T] = \begin{bmatrix} \overbrace{\mathrm{Cov}[Y_1, Y_1]}^{\mathrm{Var}[Y_1] \,=} & \cdots & \mathrm{Cov}[Y_1, Y_m] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[Y_m, Y_1] & \cdots & \underbrace{\mathrm{Cov}[Y_m, Y_m]}_{= \,\mathrm{Var}[Y_m]} \end{bmatrix}$$

→ The cross-covariance matrix $\Sigma_{\underline{X},\underline{Y}}$ between $\underline{X}$ and $\underline{Y}$ is

$$\Sigma_{\underline{X},\underline{Y}} = \mathbb{E}[(\underline{X} - \mathbb{E}[\underline{X}])(\underline{Y} - \mathbb{E}[\underline{Y}])^T] = \begin{bmatrix} \mathrm{Cov}[X_1, Y_1] & \cdots & \mathrm{Cov}[X_1, Y_m] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_n, Y_1] & \cdots & \mathrm{Cov}[X_n, Y_m] \end{bmatrix}$$

- **Vector LLSE Estimator Properties:**

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

→ The error of the vector LLSE estimator is orthogonal to any linear function of the observation:

$$\mathbb{E}\left[(\underline{X} - \hat{\underline{x}}_{LLSE}(\underline{Y}))(A\underline{Y} + \underline{b})^T\right] = \mathbf{0} \quad \leftarrow \text{all zeros matrix}$$

→ If the random vector $\begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix}$ is jointly Gaussian, the MMSE and LLSE estimators are equal $\hat{\underline{x}}_{MMSE}(\underline{y}) = \hat{\underline{x}}_{LLSE}(\underline{y})$.

→ The LLSE estimator can be derived from the unbiasedness and orthogonality properties. (See lecture notes for details.)

→ We only need first- and second-order statistics to implement this estimator. These are relatively easy to collect in practice.

- The vector LLSE estimator is frequently applied to real datasets. In statistics, it is often referred to as multivariate regression.

→ Dataset: $(\underline{X}_1, \underline{Y}_1), (\underline{X}_2, \underline{Y}_2), \ldots, (\underline{X}_n, \underline{Y}_n)$

→ Estimate mean vectors, covariance and cross-covariance matrices.

### Sample Mean Vectors

$$\hat{\underline{\mu}}_{\underline{X}} = \frac{1}{n}\sum_{i=1}^{n} \underline{X}_i \qquad \hat{\underline{\mu}}_{\underline{Y}} = \frac{1}{n}\sum_{i=1}^{n} \underline{Y}_i$$

### Sample Covariance Matrix

$$\hat{\Sigma}_{\underline{Y}} = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{Y}_i - \hat{\underline{\mu}}_{\underline{Y}})(\underline{Y}_i - \hat{\underline{\mu}}_{\underline{Y}})^T$$

### Sample Cross-Covariance Matrix

$$\hat{\Sigma}_{\underline{X},\underline{Y}} = \frac{1}{n-1}\sum_{i=1}^{n} (\underline{X}_i - \hat{\underline{\mu}}_{\underline{X}})(\underline{Y}_i - \hat{\underline{\mu}}_{\underline{Y}})^T$$

### Multivariate Regression

$$\hat{\underline{x}}(\underline{y}) = \hat{\underline{\mu}}_{\underline{X}} + \hat{\Sigma}_{\underline{X},\underline{Y}} \hat{\Sigma}_{\underline{Y}}^{-1} (\underline{y} - \hat{\underline{\mu}}_{\underline{Y}})$$

### Mean-Squared Error

$$\widehat{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\underline{X}_i - \hat{\underline{x}}(\underline{Y}_i))^T (\underline{X}_i - \hat{\underline{x}}(\underline{Y}_i))$$