Random Vectors

- In principle, going from a pair of random variables X and Y to n random variables $X_1, X_2, \ldots, X_n$ is easy.
  - → The joint cumulative distribution function (CDF) is

$$\boxed{F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = \mathbb{P}[X_1 \le x_1,\ X_2 \le x_2,\ \ldots,\ X_n \le x_n]}$$

  - → For discrete random variables, the joint probability mass function (PMF) is

$$\boxed{P_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = \mathbb{P}[X_1 = x_1,\ X_2 = x_2,\ \ldots,\ X_n = x_n]}$$

  - → For continuous random variables, we have a joint probability density function (PDF) satisfying

$$\boxed{F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{X_1,X_2,\ldots,X_n}(u_1,u_2,\ldots,u_n)\, du_1\, du_2 \cdots du_n}$$

- The usual PMF/PDF properties hold:
  - → Non-negativity: $p_{X_1,\dots,X_n}(x_1,\dots,x_n) \geq 0$

    $$f_{X_1,\dots,X_n}(x_1,\dots,x_n) \geq 0$$
  - → Normalization: $\sum_{x_1 \in R_{X_1}} \cdots \sum_{x_n \in R_{X_n}} p_{X_1,\dots,X_n}(x_1,\dots,x_n) = 1$

    $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 \cdots dx_n = 1$$
  - → Additivity: $\mathbb{P}[\{(X_1,\dots,X_n) \in B\}] = \sum_{(x_1,\dots,x_n)\in B} p_{X_1,\dots,X_n}(x_1,\dots,x_n)$

    $$\mathbb{P}[\{(X_1,\dots,X_n) \in B\}] = \int \cdots \int_B f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 \cdots dx_n$$
- $X_1,\dots,X_n$ are independent if $p_{X_1,\dots,X_n}(x_1,\dots,x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$

  $$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

- The expected value $\mathbb{E}[g(X_1, \dots, X_n)]$ of a function $g(x_1, \dots, x_n)$ of n random variables is

$$\mathbb{E}[g(X_1,\dots,X_n)] = \sum_{x_1 \in R_{X_1}} \cdots \sum_{x_n \in R_{X_n}} g(x_1,\dots,x_n)\, P_{X_1,\dots,X_n}(x_1,\dots,x_n)$$

(Discrete)

$$\mathbb{E}[g(X_1,\dots,X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1,\dots,x_n)\, f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 \cdots dx_n$$

(Continuous)

- Linearity of Expectation: For any functions $g_1(x_1, \dots, x_n), \dots, g_m(x_1, \dots, x_n)$ and constants $a_1, \dots, a_m$,

$$\mathbb{E}[a_1 g_1(X_1,\dots,X_n) + \cdots + a_m g_m(X_1,\dots,X_n)]$$
$$= a_1 \mathbb{E}[g_1(X_1,\dots,X_n)] + \cdots + a_m \mathbb{E}[g_m(X_1,\dots,X_n)]$$

- **Conditioning** also generalizes naturally. Say we want the conditional distribution of $X_1, \ldots, X_m$ given $X_{m+1}, \ldots, X_n$. We just divide the joint distribution by the marginal distribution:

→ The conditional PMF of $X_1, \ldots, X_m$ given $X_{m+1}, \ldots, X_n$ is

$$p_{X_1,\ldots,X_m | X_{m+1},\ldots,X_n}(x_1,\ldots,x_m | x_{m+1},\ldots,x_n) = \begin{cases} \dfrac{p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{p_{X_{m+1},\ldots,X_n}(x_{m+1},\ldots,x_n)} & \text{for } (x_1,\ldots,x_n) \text{ in } R_{X_1,\ldots,X_n} \\ 0 & \text{otherwise} \end{cases}$$

→ The conditional PDF of $X_1, \ldots, X_m$ given $X_{m+1}, \ldots, X_n$ is

$$f_{X_1,\ldots,X_m | X_{m+1},\ldots,X_n}(x_1,\ldots,x_m | x_{m+1},\ldots,x_n) = \begin{cases} \dfrac{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{f_{X_{m+1},\ldots,X_n}(x_{m+1},\ldots,x_n)} & \text{for } (x_1,\ldots,x_n) \text{ in } R_{X_1,\ldots,X_n} \\ 0 & \text{otherwise} \end{cases}$$

→ The conditional expected value $\mathbb{E}[g(X_1,\ldots,X_n) | X_{m+1} = x_{m+1}, \ldots, X_n = x_n]$ is

Discrete: $$\sum_{x_1 \in R_{X_1}} \cdots \sum_{x_m \in R_{X_m}} g(x_1,\ldots,x_m,x_{m+1},\ldots,x_n)\, p_{X_1,\ldots,X_m | X_{m+1},\ldots,X_n}(x_1,\ldots,x_m | x_{m+1},\ldots,x_n)$$

Continuous: $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1,\ldots,x_m,x_{m+1},\ldots,x_n)\, f_{X_1,\ldots,X_m | X_{m+1},\ldots,X_n}(x_1,\ldots,x_m | x_{m+1},\ldots,x_n)\, dx_1 \cdots dx_m$$

- The main issue with $n > 2$ random variables is that working out n-dimensional sums and integrals is hard.

- However, in some cases, we can get by with only working out first- and second-order statistics, such as the mean and covariance.

- For n random variables $X_1, \dots, X_n$, we would need
  - → n means $\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]$
  - → n^2 covariances $\mathrm{Cov}[X_i, X_j] = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$ for $i, j \in \{1, \dots, n\}$
    (Note that $\mathrm{Cov}[X_i, X_i] = \mathrm{Var}[X_i]$.)

These are just one- and two-dimensional calculations (and relatively easy to learn from data).

- Why are the means and covariances useful?
  - → One important reason: Easy to work out how means and covariances evolve under linear transformations.
  - → Another reason: Jointly Gaussian random variables are fully specified by their means and covariances.
- To explore these ideas, we need vectors and matrices.
- A random vector $\underline{X}$ is a column vector whose entries are random variables.
  - → Notation: $\underbrace{\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}}_{\text{random vector}}$ $\underbrace{\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\text{value } \underline{X} \text{ takes}}$

$$\underbrace{P_{\underline{X}}(\underline{x}) = P_{X_1,\dots,X_n}(x_1,\dots,x_n)}_{\text{Joint PMF}}$$

$$\underbrace{f_{\underline{X}}(\underline{x}) = f_{X_1,\dots,X_n}(x_1,\dots,x_n)}_{\text{Joint PDF}}$$

- We can organize the means and covariances into vectors and matrices.

- The mean vector $\underline{\mu}_{\underline{x}}$ of a random vector $\underline{X}$ is the column vector whose entries are the expected values of the corresponding entries of $\underline{X}$:

$$\boxed{\underline{\mu}_{\underline{x}} = \mathbb{E}[\underline{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}}$$

In other words, to take the expectation of a random vector, just take the expectation of each entry.

- Linearity of Expectation: For any (constant) matrix $\mathbf{A}$ and vector $\underline{b}$,

$$\boxed{\mathbb{E}[\mathbf{A}\underline{X} + \underline{b}] = \mathbf{A}\,\mathbb{E}[\underline{X}] + \underline{b}}$$

- The covariance matrix $\Sigma_{\underline{X}}$ of a random vector $\underline{X}$ is

$$\boxed{\Sigma_{\underline{X}} = \mathbb{E}\left[(\underline{X} - \mathbb{E}[\underline{X}])(\underline{X} - \mathbb{E}[\underline{X}])^T\right]}$$

$$= \mathbb{E}\left[\begin{bmatrix} X_1 - \mathbb{E}[X_1] \\ X_2 - \mathbb{E}[X_2] \\ \vdots \\ X_n - \mathbb{E}[X_n] \end{bmatrix}\begin{bmatrix} X_1 - \mathbb{E}[X_1] & X_2 - \mathbb{E}[X_2] & \cdots & X_n - \mathbb{E}[X_n] \end{bmatrix}\right]$$

$(i,j)^{th}$ entry contains $\mathrm{Cov}[X_i, X_j]$

$$= \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_2, X_1] & \mathrm{Var}[X_2] & \cdots & \mathrm{Cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_n, X_1] & \mathrm{Cov}[X_n, X_2] & \cdots & \mathrm{Var}[X_n] \end{bmatrix}$$

- Alternate Covariance Matrix Equation:

$$\Sigma_{\underline{x}} = \mathbb{E}[\underline{X}\,\underline{X}^T] - \mathbb{E}[\underline{X}](\mathbb{E}[\underline{X}])^T$$

- Covariance Matrix Properties: The covariance matrix $\Sigma_{\underline{x}}$

→ is symmetric $\Sigma_{\underline{x}} = \Sigma_{\underline{x}}^T$ (since $\mathrm{Cov}[X_i, X_j] = \mathrm{Cov}[X_j, X_i]$)

→ is positive semi-definite, $\underline{a}^T \Sigma_{\underline{x}} \underline{a} \geq 0$ for any vector $\underline{a}$

→ has all real, non-negative eigenvalues

→ has n distinct eigenvectors, each perpendicular to the others

These properties will be useful later in the course.

Covariance Matrix after a Linear Transformation:

Let $\underline{X}$ be a random vector with covariance matrix $\Sigma_{\underline{X}}$ and let $\underline{Y} = A\underline{X} + \underline{b}$. Then, the covariance matrix of $\underline{Y}$ is $\Sigma_{\underline{Y}} = A\Sigma_{\underline{X}}A^T$.

Why?

$$\Sigma_{\underline{Y}} = \mathbb{E}\left[(\underline{Y} - \mathbb{E}[\underline{Y}])(\underline{Y} - \mathbb{E}[\underline{Y}])^T\right]$$

$$= \mathbb{E}\left[(A\underline{X} + \underline{b} - \mathbb{E}[A\underline{X} + \underline{b}])(A\underline{X} + \underline{b} - \mathbb{E}[A\underline{X} + \underline{b}])^T\right]$$

Linearity of Expectation

$$= \mathbb{E}\left[(A\underline{X} + \cancel{\underline{b}} - A\mathbb{E}[\underline{X}] - \cancel{\underline{b}})(A\underline{X} + \cancel{\underline{b}} - A\mathbb{E}[\underline{X}] - \cancel{\underline{b}})^T\right]$$

$$= \mathbb{E}\left[A(\underline{X} - \mathbb{E}[\underline{X}])(\underline{X} - \mathbb{E}[\underline{X}])^T A^T\right]$$

$(AC)^T = C^T A^T$

Linearity of Expectation (twice)

$$= A\,\mathbb{E}\left[(\underline{X} - \mathbb{E}[\underline{X}])(\underline{X} - \mathbb{E}[\underline{X}])^T\right]A^T$$

$$= A\Sigma_{\underline{X}}A^T$$

- **Example:** $\underline{X}$ is a random vector with mean vector $\underline{\mu_X} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and covariance matrix $\Sigma_{\underline{X}} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

  Let $\underline{Y} = \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix} \underline{X} + \begin{bmatrix} 4 \\ -3 \end{bmatrix}$. Determine $\underline{\mu_Y}$ and $\Sigma_{\underline{Y}}$.

Linearity of Expectation: $\mathbb{E}[A\underline{X} + \underline{b}] = A\,\mathbb{E}[\underline{X}] + \underline{b}$

$$\underline{\mu_Y} = \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 4 \\ -3 \end{bmatrix} = \begin{bmatrix} -1\cdot 1 + 1\cdot(-1) \\ 3\cdot 1 + 2\cdot(-1) \end{bmatrix} + \begin{bmatrix} 4 \\ -3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ -3 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

Covariance of a Linear Transformation: $\Sigma_{\underline{Y}} = A\,\Sigma_{\underline{X}}\,A^T$

$$\Sigma_{\underline{Y}} = \begin{bmatrix} -1 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} -1\cdot 2 + 1\cdot 1 & -1\cdot 1 + 1\cdot 2 \\ 3\cdot 2 + 2\cdot 1 & 3\cdot 1 + 2\cdot 2 \end{bmatrix}\begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & 1 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} -1\cdot(-1) + 1\cdot 1 & -1\cdot 3 + 1\cdot 2 \\ 8\cdot(-1) + 7\cdot 1 & 8\cdot 3 + 7\cdot 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 38 \end{bmatrix}$$