5.1 Covariance and Correlation

- The **covariance** of random variables X and Y is
$$\mathsf{Cov}[X,Y] = \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\big].$$
- Another useful formula is $\mathsf{Cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.
- Intuition: Captures the (average) linear relationship between $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$.
  - If $\mathsf{Cov}[X,Y] > 0$, then $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ tend to have the same sign and a line with positive slope will fit the data better.
  - If $\mathsf{Cov}[X,Y] < 0$, then $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ tend to have the opposite sign and a line with negative slope will fit the data better.
- The covariance satisfies the following basic properties:
  - $\mathsf{Cov}[X,Y] = \mathsf{Cov}[Y,X]$
  - $\mathsf{Cov}[X,X] = \mathsf{Var}[X]$
  - $\mathsf{Cov}[X,a] = 0$ for any number a.
  - **Variance of Sums:** $\mathsf{Var}[X+Y] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\mathsf{Cov}[X,Y]$.
  - **Variance of Linear Functions:** $\mathsf{Var}[aX+bY+c] = a^2\,\mathsf{Var}[X] + b^2\,\mathsf{Var}[Y] + 2ab\,\mathsf{Cov}[X,Y]$.
  - **Covariance of Linear Functions:**
    $\mathsf{Cov}[aX+bY+c, dX+eY+f] = ad\,\mathsf{Var}[X] + be\,\mathsf{Var}[Y] + (ae+bd)\mathsf{Cov}[X,Y]$.
- The **correlation coefficient** is $\rho_{X,Y} = \dfrac{\mathsf{Cov}[X,Y]}{\sqrt{\mathsf{Var}[X]\mathsf{Var}[Y]}}$.
- Intuition: The correlation coefficient as a "scale-invariant" version of the covariance. The closer $|\rho_{X,Y}|$ is to 1, the better a line explains the relationship between X and Y.
- The correlation coefficient satisfies the following basic properties:
  - $-1 \leq \rho_{X,Y} \leq 1$.
  - $\rho_{X,Y} = 1$ if and only if $X = aY + b$ for some $a > 0$ and any b.
  - $\rho_{X,Y} = -1$ if and only if $X = aY + b$ for some $a < 0$ and any b.
  - **Correlation Coefficient of Linear Functions:** If $U = aX + b$ and $V = cY + d$, then
$$\rho_{U,V} = \text{sign}(ac)\rho_{X,Y} \quad \text{where} \quad \text{sign}(z) = \begin{cases} +1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases}$$
- Two random variables X and Y are **uncorrelated** if $\mathsf{Cov}[X,Y] = 0$ (or $\rho_{X,Y} = 0$).
  - If X and Y are uncorrelated, we have that
    - $\mathsf{Var}[X+Y] = \mathsf{Var}[X] + \mathsf{Var}[Y]$
    - $\mathsf{Var}[aX+bY+c] = a^2\,\mathsf{Var}[X] + b^2\,\mathsf{Var}[Y]$
    - $\mathsf{Cov}[aX+bY+c, dX+eY+f] = ad\,\mathsf{Var}[X] + be\,\mathsf{Var}[Y]$
    - $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
  - Independence implies uncorrelatedness but uncorrelatedness does not imply independence.

5.2 Jointly Gaussian Random Variables

- U and V are called **independent, standard Gaussian random variables** if they are independent Gaussian$(0,1)$ random variables. In this case, the joint PDF is

$$f_{U,V}(u,v) = \frac{1}{2\pi} \exp\Big(-\frac{1}{2}(u^2+v^2)\Big).$$

- X and Y are **jointly Gaussian random variables** if they can be expressed as linear functions of independent, standard Gaussian random variables

$$X = aU + bV + c \qquad Y = dU + eV + f \ .$$

  However, this representation is usually left implicit, and the joint Gaussian distribution of X and Y is specified by 5 parameters:

  - Means: $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$
  - Variances: $\sigma_X^2 = \mathsf{Var}[X]$, $\sigma_Y^2 = \mathsf{Var}[Y]$
  - Covariance: $\mathsf{Cov}[X,Y]$ **or** Correlation Coefficient: $\rho_{X,Y}$.

- The joint PDF is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \exp\left(-\frac{1}{2(1-\rho_{X,Y}^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho_{X,Y}\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right)$$

- Jointly Gaussian random variables X and Y satisfy the following properties:

  - **Linear functions are Gaussian:** If $W = aX + bY + c$ and $Z = dX + eY + f$, then W and Z are jointly Gaussian with parameters that be determined via the linearity of expectation and the variance and covariance of linear functions.
  - **Marginal PDFs are Gaussian:** X is Gaussian(μ_X, σ_X) and Y is Gaussian(μ_Y, σ_Y).
  - **Uncorrelated implies Independence:** X and Y are uncorrelated ($\rho_{X,Y} = 0$) if and only if X and Y are independent.
  - **Conditional Expected Value:** $$\mathbb{E}[X|Y=y] = \mu_X + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y) = \mu_X + \frac{\mathsf{Cov}[X,Y]}{\mathsf{Var}[Y]}(y-\mu_Y)$$
  - **Conditional Variance:** $\sigma_{X|Y}^2 = \mathsf{Var}[X|Y=y] = (1-\rho_{X,Y}^2)\sigma_X^2$.
  - **Conditional PDF is Gaussian:** The conditional PDF $f_{X|Y}(x|y)$ of X given Y is Gaussian$(\mathbb{E}[X|Y=y],\ \sigma_{X|Y}^2)$.

5.3 More than Two Random Variables

- All of the concepts from pairs of random variables generalize to n random variables $X_1, \dots, X_n$.

- The **joint cumulative distribution function (CDF)** is

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = \mathbb{P}[X_1 \le x_1, \dots, X_n \le x_n] \ .$$

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

$$P_{X_1,\dots,X_n}(x_1,\dots,x_n) = \mathbb{P}[X_1 = x_1,\dots,X_n = x_n]$$

and the **range** is $R_{X_1,\dots,X_n} = \{(x_1,\dots,x_n) : P_{X_1,\dots,X_n}(x_1,\dots,x_n) > 0\}$.

- For continuous random variables, the **joint probability density function (PDF)** is

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = \frac{\partial^n F_{X_1,\dots,X_n}(x_1,\dots,x_n)}{\partial x_1 \cdots \partial x_n}$$

and the **range** is $R_{X_1,\dots,X_n} = \{(x_1,\dots,x_n) : f_{X_1,\dots,X_n}(x_1,\dots,x_n) > 0\}$.

- The basic PMF/PDF properties apply:
  - **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:** $\displaystyle\sum_{x_1 \in R_{X_1}} \cdots \sum_{x_n \in R_{X_n}} P_{X_1,\dots,X_n}(x_1,\dots,x_n) = 1$
  $\displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 \cdots dx_n = 1$
  - **Probability of an event:**

$$\mathbb{P}\big[(X_1,\dots,X_n) \in B\big] = \begin{cases} \displaystyle\sum_{(x_1,\dots,x_n)\in B} P_{X_1,\dots,X_n}(x_1,\dots,x_n) & \text{Discrete} \\ \displaystyle\int \cdots \int_B f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 \dots dx_n & \text{Continuous} \end{cases}$$

- $X_1,\dots,X_n$ are **independent** if and only if the joint PMF/PDF factors into the product of the marginal PMFs/PDFs: $P_{X_1,\dots,X_n}(x_1,\dots,x_n) = P_{X_1}(x_1)\cdots P_{X_n}(x_n)$
$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$

- To obtain a **marginal** PMF/PDF of a subset $X_1,\dots,X_m$ of the random variables, we sum/integrate over the undesired variables $X_{m+1},\dots,X_n$:

$$P_{X_1,\dots,X_m}(x_1,\dots,x_m) = \sum_{x_{m+1}\in R_{X_{m+1}}} \cdots \sum_{x_n \in R_{X_n}} P_{X_1,\dots,X_n}(x_1,\dots,x_n)$$

$$f_{X_1,\dots,X_m}(x_1,\dots,x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_{m+1}\cdots dx_n$$

- The **expected value of a function** is

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

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

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

$$\begin{aligned} &\mathbb{E}\big[a_1 g_1(X_1,\dots,X_n) \;+\; \cdots \;+\; a_m g_m(X_1,\dots,X_n)\big] \\ &= a_1\mathbb{E}[g_1(X_1,\dots,X_n)] \;+\; \cdots \;+\; a_m\mathbb{E}[g_m(X_1,\dots,X_n)] \end{aligned}$$

- The **conditional PMF** of $X_1, \dots, X_m$ given $X_{m+1}, \dots, X_n$ is

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

- The **conditional PDF** of $X_1, \dots, X_m$ given $X_{m+1}, \dots, X_n$ is

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

- The **conditional expected value** is

$$\mathbb{E}\big[g(X_1,\dots,X_n) \,\big|\, X_{m+1} = x_{m+1},\dots,X_n = x_n\big]$$
$$= \begin{cases} \displaystyle\sum_{x_1 \in R_{X_1}} \cdots \sum_{x_m \in R_{X_m}} g(x_1,\dots,x_n)\, P_{X_1,\dots,X_m|X_{m+1},\dots,X_n}(x_1,\dots,x_m|x_{m+1},\dots,x_n) & \text{Discrete} \\ \displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1,\dots,x_n)\, f_{X_1,\dots,X_m|X_{m+1},\dots,X_n}(x_1,\dots,x_m|x_{m+1},\dots,x_n)\, dx_1 \cdots dx_m & \text{Continuous} \end{cases}$$

5.4 Random Vectors

- A **random vector** $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ is a vector whose entries $X_1, \dots, X_n$ are random variables.
  - If the entries are discrete random variables, the random vector has a joint PMF $P_{\underline{X}}(\underline{x}) = P_{X_1,\dots,X_n}(x_1,\dots,x_n)$.
  - If the entries are continuous random variables, the random vector has a joint PDF $f_{\underline{X}}(\underline{x}) = f_{X_1,\dots,X_n}(x_1,\dots,x_n)$.

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

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

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

- The **covariance matrix** $\mathbf{\Sigma}_{\underline{X}}$ is a matrix whose $(i,j)^{\text{th}}$ entry is the covariance between the i^{th} and j^{th} entry of the vector,

$$\mathbf{\Sigma}_{\underline{X}} = \mathbb{E}\big[(\underline{X} - \mathbb{E}[\underline{X}])(\underline{X} - \mathbb{E}[\underline{X}])^{\mathsf{T}}\big] = \begin{bmatrix} \mathsf{Cov}[X_1, X_1] & \cdots & \mathsf{Cov}[X_1, X_n] \\ \vdots & & \vdots \\ \mathsf{Cov}[X_n, X_1] & \cdots & \mathsf{Cov}[X_n, X_n] \end{bmatrix}$$

- Another useful formula is $\mathbf{\Sigma}_{\underline{X}} = \mathbb{E}\big[\underline{X}\,\underline{X}^{\mathsf{T}}\big] - \mathbb{E}\big[\underline{X}\big]\big(\mathbb{E}\big[\underline{X}\big]\big)^{\mathsf{T}}$

- **Covariance of a Linear Transform:** If $\underline{Y} = \mathbf{A}\underline{X} + \underline{b}$, then $\mathbf{\Sigma}_{\underline{Y}} = \mathbf{A}\mathbf{\Sigma}_{\underline{X}}\mathbf{A}^{\mathsf{T}}$.

- The covariance matrix $\mathbf{\Sigma}_{\underline{X}}$ satisfies the following properties:
  - Symmetry: $\mathbf{\Sigma}_{\underline{X}} = \mathbf{\Sigma}_{\underline{X}}^{\mathsf{T}}$
  - Positive Semi-Definite: $\underline{a}^{\mathsf{T}}\mathbf{\Sigma}_{\underline{X}}\underline{a} \geq 0$
  - n real, non-negative eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ and n real eigenvectors $\underline{v}_1, \dots, \underline{v}_n \in \mathbb{R}^n$ that are orthonormal, $\underline{v}_i^{\mathsf{T}}\underline{v}_j = 0$ for $i \neq j$ and $\underline{v}_i^{\mathsf{T}}\underline{v}_i = 1$ for $i = 1, \dots, n$.
  - Collecting the eigenvalues and the eigenvectors into matrices
$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \qquad \mathbf{V} = \begin{bmatrix} | & & | \\ \underline{v}_1 & \cdots & \underline{v}_n \\ | & & | \end{bmatrix},$$
we have that $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ is a diagonal matrix with real, non-negative entries and $\mathbf{V} \in \mathbb{R}^{n\times n}$ is a real, orthogonal matrix, $\mathbf{V}\mathbf{V}^{\mathsf{T}} = \mathbf{V}^{\mathsf{T}}\mathbf{V} = \mathbf{I}$. The eigendecomposition of the covariance matrix is $\mathbf{\Sigma}_{\underline{X}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{T}}$.

5.5 Gaussian Vectors

- A **standard Gaussian vector** is a random vector $\underline{Z}$ whose entries $Z_1, \dots, Z_n$ are independent Gaussian$(0, 1)$ random variables.
- A **(jointly) Gaussian vector** is a random vector $\underline{X}$ that can be written as a linear transform $\underline{X} = \mathbf{A}\underline{Z} + \underline{b}$ of a standard Gaussian vector $\underline{Z}$. However, this representation is usually left implicit, and the distribution of a Gaussian vector $\underline{X}$ is specified by its mean vector $\underline{\mu}_{\underline{X}}$ and covariance matrix $\mathbf{\Sigma}_{\underline{X}}$.
- Shorthand notation: We often write $\underline{X} \sim \mathcal{N}(\underline{\mu}_{\underline{X}}, \mathbf{\Sigma}_{\underline{X}})$ to mean that $\underline{X}$ is a Gaussian vector with mean vector $\underline{\mu}_{\underline{X}}$ and covariance matrix $\mathbf{\Sigma}_{\underline{X}}$.
- A Gaussian vector $\underline{X} \sim \mathcal{N}(\underline{\mu}_{\underline{X}}, \mathbf{\Sigma}_{\underline{X}})$ satisfies the following properties:
  - The entries of $\underline{X}$ are independent if and only if $\mathbf{\Sigma}_{\underline{X}}$ is a diagonal matrix.
  - For any choice of vector $\underline{a} \in \mathbb{R}^n$, $\underline{a}^{\mathsf{T}}\underline{X}$ is a scalar Gaussian random variable.
  - If $\mathbf{\Sigma}_{\underline{X}}$ is invertible, the joint PDF of $\underline{X} \sim \mathcal{N}(\underline{\mu}_{\underline{X}}, \mathbf{\Sigma}_{\underline{X}})$ is
$$f_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{\Sigma}_{\underline{X}})}} \exp\left(-\frac{1}{2}(\underline{x} - \underline{\mu}_{\underline{X}})^{\mathsf{T}}\mathbf{\Sigma}_{\underline{X}}^{-1}(\underline{x} - \underline{\mu}_{\underline{X}})\right)$$
  - **Linear transformations of Gaussian vectors are themselves Gaussian vectors:** If $\underline{Y} = \mathbf{B}\underline{X} + \underline{c}$, then $\underline{Y} \sim \mathcal{N}(\mathbf{B}\underline{\mu}_{\underline{X}} + \underline{c}, \mathbf{B}\mathbf{\Sigma}_{\underline{X}}\mathbf{B}^{\mathsf{T}})$.
  - Let $\underline{Y}$ be a Gaussian vector with mean vector $\underline{\mu}_{\underline{Y}}$ and covariance matrix $\mathbf{\Sigma}_{\underline{Y}}$, and assume that $\begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix}$ concatenated together is also a Gaussian vector with mean vector $\begin{bmatrix} \underline{\mu}_{\underline{X}} \\ \underline{\mu}_{\underline{Y}} \end{bmatrix}$ and covariance matrix $\begin{bmatrix} \mathbf{\Sigma}_{\underline{X}} & \mathbf{\Sigma}_{\underline{X},\underline{Y}} \\ \mathbf{\Sigma}_{\underline{X},\underline{Y}}^{\mathsf{T}} & \mathbf{\Sigma}_{\underline{Y}} \end{bmatrix}$ where $\mathbf{\Sigma}_{\underline{X},\underline{Y}} = \mathbb{E}[(\underline{X} - \underline{\mu}_{\underline{X}})(\underline{Y} - \underline{\mu}_{\underline{Y}})^{\mathsf{T}}]$ is sometimes called the cross-covariance matrix. Then, the conditional PDF of $\underline{X}$ given $\underline{Y}$ is Gaussian with mean vector $\mathbb{E}[\underline{X}|\underline{Y} = \underline{y}]$ and covariance matrix $\mathbf{\Sigma}_{\underline{X}|\underline{Y}}$ where
$$\mathbb{E}[\underline{X}|\underline{Y} = \underline{y}] = \underline{\mu}_{\underline{X}} + \mathbf{\Sigma}_{\underline{X},\underline{Y}}\mathbf{\Sigma}_{\underline{X}}^{-1}(\underline{y} - \underline{\mu}_{\underline{Y}})$$
$$\mathbf{\Sigma}_{\underline{X}|\underline{Y}} = \mathbf{\Sigma}_{\underline{X}} - \mathbf{\Sigma}_{\underline{X},\underline{Y}}\mathbf{\Sigma}_{\underline{Y}}^{-1}\mathbf{\Sigma}_{\underline{X},\underline{Y}}^{\mathsf{T}}$$