4.1 Pairs of Random Variables

- Formally, **multiple random variables** are the result of a mapping that assigns multiple real numbers to outcomes in the sample space.
- Intuitively, we can think of multiple random variables as the observations from an experiment that simultaneously produces two or more numbers.
- Multiple random variables are denoted by capital letters and their values by lowercase letters.
  - n random variables are often denoted by $X_1, X_2, \ldots, X_n$ and their values by $x_1, x_2, \ldots, x_n$
  - a pair of random variables ($n = 2$) is often denoted by X, Y and their values by x, y
- The relationship between multiple random variables is more general than a function. A function maps one number to another number whereas, for a given value of X, the random variable Y may randomly take one of several values.
- Most of the basic concepts are well-captured by the special case of pairs of random variables, which we focus on below.

4.1.1 Joint Cumulative Distribution Function

- The **joint cumulative distribution function (CDF)** returns the probability that the random variables X and Y are less than or equal to the values x and y, respectively:

$$F_{X,Y}(x,y) = \mathbb{P}[X \leq x,\ Y \leq y] = \mathbb{P}\big[\{X \leq x\} \cap \{Y \leq y\}\big].$$

- Unifies discrete and continuous random variables.
- The joint CDF satisfies the following basic properties:
  - **Non-negativity:** $F_{X,Y}(x,y) \geq 0$.
  - **Normalization:** $\lim\limits_{x,y\to\infty} F_{X,Y}(x,y) = 1$.
  - **Non-decreasing:** For any $x \leq \tilde{x}$ and $y \leq \tilde{y}$, $F_{X,Y}(x,y) \leq F_{X,Y}(\tilde{x},\tilde{y})$.
  - **Marginalization:** $\lim\limits_{y\to\infty} F_{X,Y}(x,y) = F_X(x)$ and $\lim\limits_{x\to\infty} F_{X,Y}(x,y) = F_Y(y)$.

4.2 Pairs of Discrete Random Variables

- A pair of random variables X, Y is discrete if X and Y are discrete random variables.

4.2.1 Joint Probability Mass Function

- The **joint probability mass function (PMF)** of a pair of discrete random variables X and Y is

$$P_{X,Y}(x,y) = \mathbb{P}[X = x,\ Y = x] = \mathbb{P}\big[\{X = x\} \cap \{Y = y\}\big].$$

- The **range** $R_{X,Y}$ of a pair of discrete random variables is the set of all possible pairs of values,

$$R_{X,Y} = \{(x,y) : P_{X,Y}(x,y) > 0\}.$$

- The joint PMF satisfies the following basic properties:

- **Non-negativity:** $P_{X,Y}(x,y) \geq 0$.
- **Normalization:** $\displaystyle\sum_{(x,y)\in R_{X,Y}} P_{X,Y}(x,y) = 1$.
- **Additivity:** $\mathbb{P}\big[(X,Y) \in B\big] = \displaystyle\sum_{(x,y)\in B} P_{X,Y}(x,y)$.

4.2.2 Marginal PMF

- The **marginal PMF** $P_X(x)$ is just the PMF of X. Similarly, the marginal PMF $P_Y(y)$ is just the PMF Y.
- A marginal PMF can be obtained by summing the joint PMF over the undesired variable:

$$P_X(x) = \sum_{y\in R_Y} P_{X,Y}(x,y) \qquad P_Y(y) = \sum_{x\in R_X} P_{X,Y}(x,y)$$

4.2.3 Conditional PMF

- The **conditional PMF** gives the probability of one random variable when the other is fixed to a certain value:

$$\text{Conditional PMF of X given Y:} \quad P_{X|Y}(x|y) = \mathbb{P}[X = x|Y = y] = \begin{cases} \dfrac{P_{X,Y}(x,y)}{P_Y(y)} & (x,y) \in R_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Conditional PMF of Y given X:} \quad P_{Y|X}(y|x) = \mathbb{P}[Y = y|X = x] = \begin{cases} \dfrac{P_{X,Y}(x,y)}{P_X(x)} & (x,y) \in R_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

- The conditional PMF satisfies the following basic properties:
  - **Non-negativity:** $P_{X|Y}(x|y) \geq 0$ and $P_{Y|X}(y|x) \geq 0$ for all x and y.
  - **Normalization:** $\displaystyle\sum_{x\in R_X} P_{X|Y}(x|y) = 1$ for any y and $\displaystyle\sum_{y\in R_Y} P_{Y|X}(y|x) = 1$ for any x.
  - **Additivity:** For any event $B \subset R_X$, the probability that X falls in B given $Y = y$ is

$$\mathbb{P}[X \in B|Y = y] = \sum_{x\in B} P_{X|Y}(x|y).$$

    For any event $B \subset R_Y$, the probability that Y falls in B given $X = x$ is

$$\mathbb{P}[Y \in B|X = x] = \sum_{y\in B} P_{Y|X}(y|x).$$

- The techniques we developed for conditional probabilities also apply to conditional PMFs:
  - **Multiplication Rule:** $P_{X,Y}(x,y) = P_{X|Y}(x|y)P_Y(y) = P_{Y|X}(y|x)P_X(x)$.
  - **Law of Total Probability:** $P_X(x) = \displaystyle\sum_{y\in R_Y} P_{X|Y}(x|y)P_Y(y) \qquad P_Y(y) = \sum_{x\in R_X} P_{Y|X}(y|x)P_X(x)$.
  - **Bayes' Rule:** $P_{X|Y}(x|y) = \dfrac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)} \qquad P_{Y|X}(y|x) = \dfrac{P_{X|Y}(x|y)P_Y(y)}{P_X(x)}$.

4.3 Pairs of Continuous Random Variables

- A pair of random variables X, Y is continuous if their joint CDF is continuous and differentiable almost everywhere.

4.3.1 Joint Probability Density Function

- The **joint probability density function (PDF)** of a pair of continuous random variables X and Y is

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} & \text{if } F_{X,Y}(x,y) \text{ is differentiable at } (x,y), \\ \text{any non-negative value} & \text{otherwise.} \end{cases}$$

- The **range** $R_{X,Y}$ of a pair of continuous random variables is the set of all possible pairs of values,

$$R_{X,Y} = \big\{(x,y) : f_{X,Y}(x,y) > 0\big\}.$$

4.3.2 Marginal PDF

- The **marginal PDF** $f_X(x)$ is just the PDF of X. Similarly, the marginal PDF $f_Y(y)$ is just the PDF of Y.
- A marginal PDF can be obtained by integrating the joint PDF over the undesired variable:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx.$$

4.3.3 Conditional PDF

- The **conditional PDF** gives the probability density of one random variable when the other is fixed to a certain value

$$\text{Conditional PDF of X given Y:} \quad f_{X|Y}(x|y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & (x,y) \in R_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Conditional PDF of Y given X:} \quad f_{Y|X}(y|x) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_X(x)} & (x,y) \in R_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

- The conditional PDF satisfies the following basic properties:
  - **Non-negativity:** $f_{X|Y}(x|y) \geq 0$ and $f_{Y|X}(y|x) \geq 0$ for all x and y.
  - **Normalization:** $\displaystyle\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = 1$ for any y and $\displaystyle\int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dy = 1$ for any x.
  - **Additivity:** For any event $B \subset R_X$, the probability that X falls in B given $Y = y$ is

    $$\mathbb{P}[X \in B | Y = y] = \int_B f_{X|Y}(x|y)\, dy.$$

    For any event $B \subset R_Y$, the probability that Y falls in B given $X = x$ is

    $$\mathbb{P}[Y \in B | X = x] = \int_B f_{Y|X}(y|x)\, dx.$$

- The techniques we developed for conditional probabilities also apply to conditional PDFs:
  - **Multiplication Rule:** $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$.
  - **Law of Total Probability:** $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)\, f_X(x)\, dx$.
  - **Bayes' Rule:** $f_{X|Y}(x|y) = \dfrac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} \qquad f_{Y|X}(y|x) = \dfrac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$.

4.4 Independence of Pairs of Random Variables

- A pair of random variables X and Y are **independent** if and only if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.
- This condition is equivalent to
  - Discrete: X and Y are independent if and only if $P_{X,Y}(x,y) = P_X(x)P_Y(y)$.
  - Continuous: X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.
- Independence can also be connected to the conditional PMF or PDF:
  - Discrete: X and Y are independent if and only if
    $P_{X|Y}(x|y) = P_X(x)$ and $P_{Y|X}(y|x) = P_Y(y)$. (Suffices to check one of these.)
  - Continuous: X and Y are independent if and only if
    $f_{X|Y}(x|y) = f_X(x)$ and $f_{Y|X}(y|x) = f_Y(y)$. (Suffices to check one of these.)
- In some special cases, we can quickly rule out independence:
  - Discrete: If there is a zero entry in the joint PMF table for which neither the entire column or entire row is zero, then the random variables are not independent.
  - Continuous: If the range is not a collection of rectangles parallel to the axes (possibly of infinite extent in either or both dimensions), then the random variables are not independent.

4.5 Expected Value of a Function of Pairs of Random Variables

- The **expected value of a function** $W = g(X,Y)$ is

$$\text{Discrete:} \quad \mathbb{E}[W] = \sum_{x \in R_X} \sum_{y \in R_Y} g(x,y) P_{X,Y}(x,y)$$

$$\text{Continuous:} \quad \mathbb{E}[W] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

- **Linearity of Expectation:** For any functions $g_1(x,y), \ldots, g_n(x,y)$ and constants $a_1, \ldots, a_n$,

$$\mathbb{E}\big[a_1 g_1(X,Y) + \cdots + a_n g_n(X,Y)\big] = a_1 \mathbb{E}[g_1(X,Y)] + \cdots + a_n \mathbb{E}[g_n(X,Y)],$$

  which does not require independence and includes the special cases
  - $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
  - For any constants a, b, c, $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$.
- **Expectation of Products:** If X and Y are independent, then $\mathbb{E}[g(X)\, h(Y)] = \mathbb{E}[g(X)]\, \mathbb{E}[h(Y)]$.

4.6 Conditional Expectation

- The **conditional expected value of X given** $Y = y$ is

$$\text{Discrete:} \quad \mathbb{E}[X|Y=y] = \sum_{x \in R_X} x\, p_{X|Y}(x|y)$$
$$\text{Continuous:} \quad \mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

  - Intuition: The conditional expected value $\mathbb{E}[X|Y=y]$ is the average value of X given that $Y = y$.
  - Think of $\mathbb{E}[X|Y=y]$ as a deterministic function of the value y.
  - Think of $\mathbb{E}[X|Y]$ as a random variable, which is a particular function of the random variable Y. To see this more clearly, define $h(y) = \mathbb{E}[X|Y=y]$ to be the function we obtain from the conditional expectation, and $\mathbb{E}[X|Y] = h(Y)$.
  - If X and Y are independent, then $\mathbb{E}[X|Y=y] = \mathbb{E}[X]$.
  - **Law of Total Expectation:** $\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \mathbb{E}[X]$. This can be easier to understand with by first defining $h(y) = \mathbb{E}[X|Y=y]$, and then substituting this in to get $\mathbb{E}[h(Y)] = \mathbb{E}[X]$.
  - We can similarly define the conditional expected value of Y given $X = x$, $\mathbb{E}[Y|X=x]$.

- The **conditional expected value of a function** $g(X)$ **given** $Y = y$ is

$$\text{Discrete:} \quad \mathbb{E}[g(X)|Y=y] = \sum_{x \in R_X} g(x)\, p_{X|Y}(x|y)$$
$$\text{Continuous:} \quad \mathbb{E}[g(X)|Y=y] = \int_{-\infty}^{\infty} g(x)\, f_{X|Y}(x|y)\, dx$$

  - If X and Y are independent, then $\mathbb{E}[g(X)|Y=y] = \mathbb{E}[g(X)]$.
  - **Law of Total Expectation:** $\mathbb{E}\big[\mathbb{E}[g(X)|Y]\big] = \mathbb{E}[g(X)]$.
  - We can similarly define the conditional expected value of a function $g(Y)$ given $X = x$, $\mathbb{E}[g(Y)|X=x]$.