

# Contents

<b>1</b>	<b>Probability Models</b>	<b>1</b>
1.1	Relative Frequency . . . . .	1
1.1.1	Properties of Relative Frequencies . . . . .	1
1.2	Statistical Regularity . . . . .	1
<b>2</b>	<b>Set Theory</b>	<b>1</b>
2.1	Properties of Set Operations . . . . .	2
<b>3</b>	<b>Probability Theory</b>	<b>2</b>
3.1	Random Experiments . . . . .	2
3.2	Probability Law Corollaries . . . . .	3
3.3	Conditional Probability . . . . .	3
3.4	Event Independence . . . . .	3
<b>4</b>	<b>Counting</b>	<b>4</b>
4.1	Sampling <i>with</i> Replacement <i>with</i> Order . . . . .	4
4.2	Sampling <i>without</i> Replacement <i>with</i> Order . . . . .	4
4.2.1	Permutations . . . . .	4
4.3	Sampling <i>with</i> Replacement <i>without</i> Order . . . . .	4
4.4	Sampling <i>without</i> Replacement <i>without</i> Ordering . . . . .	4
<b>5</b>	<b>Single Discrete Random Variables</b>	<b>5</b>
5.1	Probability Mass Function (PMF) . . . . .	5
5.1.1	Properties of Probability Mass Functions . . . . .	5
5.2	Expected Value/Mean of Single Discrete Random Variable . . . . .	5
5.2.1	Properties of Expected Values . . . . .	5
5.2.2	Moments of Random Variable . . . . .	6
5.3	Variance of Single Discrete Random Variable . . . . .	6
5.4	Conditional Probability Mass Function . . . . .	6
5.5	Conditional Expected Value of Single Discrete Random Variable . . . . .	6
5.6	Conditional Variance of Single Discrete Random Variable . . . . .	6
<b>6</b>	<b>Single Continuous Random Variables</b>	<b>7</b>
6.1	Cumulative Distribution Function (CDF) . . . . .	7
6.1.1	Properties of Cumulative Distribution Functions . . . . .	7
6.1.2	Conditional Cumulative Distribution Function . . . . .	7
6.2	Probability Density Function (PDF) . . . . .	7
6.2.1	Properties of Probability Density Functions . . . . .	8
6.2.2	Conditional Probability Density Function . . . . .	8
6.3	Expected Value of Single Continuous Random Variable . . . . .	8
6.3.1	Properties of Expected Value . . . . .	8
6.4	Variance of Single Continuous Random Variable . . . . .	9
6.5	Gaussian/Normal Random Variable . . . . .	9
6.5.1	Q-Function . . . . .	9
6.6	Markov Inequality . . . . .	9
6.7	Chebychev Inequality . . . . .	10
<b>7</b>	<b>Multiple Random Variables</b>	<b>10</b>
7.1	Joint PMF . . . . .	10
7.1.1	Marginal PMF . . . . .	11
7.2	Joint CDF . . . . .	11
7.2.1	Properties of Joint Cumulative Distribution Functions . . . . .	11
7.2.2	Marginal CDF . . . . .	11
7.3	Joint PDF . . . . .	12
7.3.1	Properties of Joint Probability Density Functions . . . . .	12
7.3.2	Facts about Joint PDFs . . . . .	12
7.3.3	Marginal PDF . . . . .	12
7.4	Independence of Multiple Random Variables . . . . .	12
7.5	Expected Value of Functions with 2 Random Variables . . . . .	13

7.6	Joint Moments, Correlation, and Covariance . . . . .	13
7.6.1	Joint Moments . . . . .	13
7.6.2	Covariance . . . . .	13
7.6.3	Correlation . . . . .	13
7.7	Conditional Probability Functions . . . . .	14
7.7.1	2 Discrete Random Variables . . . . .	14
7.7.2	1 Discrete and 1 Continuous Random Variable . . . . .	14
7.7.3	2 Continuous Random Variables . . . . .	15
7.8	Conditional Expectation of Multiple Random Variables . . . . .	15
<b>8</b>	<b>Random Vectors</b>	<b>15</b>
8.1	Joint CDF of a Random Vector . . . . .	16
8.2	Joint PDF of a Random Vector . . . . .	16
8.2.1	Marginal PDF of a Random Vector . . . . .	16
8.3	Conditional Probability Functions of Random Vectors . . . . .	16
8.3.1	Discrete Random Vectors . . . . .	16
8.3.2	Mixed Random Vectors . . . . .	16
8.3.3	Continuous Random Vectors . . . . .	16
<b>9</b>	<b>Statistics</b>	<b>16</b>
9.1	Sums of Random Variables . . . . .	16
9.1.1	Means and Variances of Sums of Random Variables . . . . .	17
<b>A</b>	<b>Reference Material</b>	<b>18</b>
A.1	Trigonometry . . . . .	18
A.1.1	Trigonometric Formulas . . . . .	18
A.1.2	Euler Equivalents of Trigonometric Functions . . . . .	18
A.2	Calculus . . . . .	18
A.2.1	Fundamental Theorems of Calculus . . . . .	18

# 1 Probability Models

## 1.1 Relative Frequency

**Defn 1** (Relative Frequency). *Relative frequency* is defined in Equation (1.1):

$$f_k(n) = \frac{N_k(n)}{n} \quad (1.1)$$

- $k$  is the outcome
- $N_k(n)$  is the number of times outcome  $k$

### 1.1.1 Properties of Relative Frequencies

(i)

$$f_k(n) = \frac{N_k(n)}{n} \quad (1.2)$$

(ii)

$$0 \leq N_k(n) \leq n \quad (1.3)$$

(iii)

$$0 \leq f_k(n) \leq 1 = \frac{0}{n} \leq \frac{N_k(n)}{n} \leq \frac{n}{n} \quad (1.4)$$

(iv)

$$\sum_{k=1}^k f_k(n) = \sum_{k=1}^k \frac{N_k(n)}{n} = \frac{\sum_{k=1}^k N_k(n)}{n} = \frac{n}{n} = 1 \quad (1.5)$$

(v)

$$\sum_{k=1}^k f_k(n) = 1 \quad (1.6)$$

(vi) If events  $A$  and  $B$  are disjoint and event  $C$  is " $A$  or  $B$ ", then

$$F_C = F_A(n) + F_B(n) \quad (1.7)$$

## 1.2 Statistical Regularity

**Defn 2.** The averages obtained in long sequences of trials that lead to approximately the same value have a property called *statistical regularity*. This is defined in Equation (1.8).

$$\lim_{n \rightarrow \infty} f_k(n) = p_k \quad (1.8)$$

- $p_k$  is the probability of event  $k$  occurring

# 2 Set Theory

1. A *set* is a collection of objects, denoted by capital letters
2. Denote the *universal set*,  $U$ ; consisting of all possible objects of interest in a given setting/application
3. For any set  $A$ , we say that " $x$  is an element of  $A$ ", denoted  $x \in A$  if object  $x$  of the universal set  $U$  is contained in  $A$
4. We say that " $x$  is not an element of  $A$ ", denoted  $x \notin A$  if object  $x$  of the universal set  $U$  is not contained in  $A$
5. We say that " $A$  is a subset of  $B$ ", denoted  $A \subset B$  if every element in  $A$  also belongs to  $B$ ,  $x \in A \rightarrow x \in B$
6. The *empty set*,  $\emptyset$  is defined as the set with no elements
  - The empty set is a subset of every set
7. Sets  $A$  and  $B$  are equal if they contain the same elements. To show this:
  - (a) Enumerate the elements of each set
  - (b) Thm:  $A = B \iff A \subset B \text{ AND } B \subset A$
8. The *union of 2 sets*  $A$ ,  $B$ , denoted  $A \cup B$  is defined as the set of outcomes that are either in  $A$ , or in  $B$ , or both
9. The *intersection of 2 sets*,  $A$ ,  $B$ , denoted  $A \cap B$  is defined as the set of outcomes in  $A$  and  $B$
10. The 2 sets  $A$ ,  $B$  are said to be *disjoint or mutually exclusive* if  $A \cap B = \emptyset$
11. The *complement of a set*  $A$ , denoted  $A^C$  is defined as the set of elements of  $U$  not in  $A$ 
  - $A^C = \{x \in U | x \notin A\}$

12. *Relative complement* or *difference*, denoted  $A - B$ , is the set of elements in  $A$  that are not in  $B$

- $A - B = A \cap B^C$
- $A^C = U - A$

## 2.1 Properties of Set Operations

Set Operators are:

1. Commutative, Equation (2.1)

$$\begin{aligned} A \cup B &= B \cup A \\ A \cap B &= B \cap A \end{aligned} \tag{2.1}$$

2. Associative, Equation (2.2)

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C \\ A \cap (B \cap C) &= (A \cap B) \cap C \end{aligned} \tag{2.2}$$

3. Distributive, Equation (2.3)

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \end{aligned} \tag{2.3}$$

4. Set Operations obey De Morgan's Laws, Equation (2.4)

$$\begin{aligned} (A \cup B)^C &= A^C \cap B^C \\ (A \cap B)^C &= A^C \cup B^C \end{aligned} \tag{2.4}$$

Additionally,

**Defn 3** (Union of  $n$  Sets). The *union of  $n$  sets*  $\bigcup_{k=1}^n A_k = A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n$  is the set consisting of all elements such that  $x \in A_k$  for some  $1 \leq k \leq n$ .

- All sets need to be empty to make  $\bigcup_{k=1}^n A_k = \emptyset$

**Defn 4** (Intersection of  $n$  Sets). The *intersection of  $n$  sets*  $\bigcap_{k=1}^n A_k = A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n$  is the set consisting of all elements such that  $x \in a_k$  for all  $1 \leq k \leq n$

- Just one set needs to be empty to make  $\bigcap_{k=1}^n A_k = \emptyset$

## 3 Probability Theory

There are 3 main components to Probability Theory.

1. Set Theory
2. Probability Law Corollaries
3. Conditional Probability and Event Independence

### 3.1 Random Experiments

**Defn 5** (Random Experiment). A *random experiment* is an experiment whose outcome varies in an unpredictable fashion when performed under the same conditions.

**Defn 6** (Sample Space). A *sample space*,  $S$  of a random experiment is the set of all possible experiments.

**Defn 7** (Outcome/Sample Point). An *outcome*, or *sample point* of a random experiment is a result that cannot be decomposed into other results.

**Defn 8** (Event). An *event* corresponds to a subset of the sample space. We say an event occurs if and only if (iff) the outcome of the experiment is in the subset representing the event.

**Defn 9** (Event Classes). An *event class*  $\mathcal{F}$  is the collection of the all the events' sets.  $\mathcal{F}$  should be closed under unions, intersections, and complements.

- For  $S$  finite, or countably infinite, then we can let  $\mathcal{F}$  be all subsets of  $S$ .

- For  $S$  uncountably infinite, instead we can let  $\mathcal{F}$  consist of the subsets that can be obtained as countable unions and intersections of some sets of  $\mathcal{F}$ .

**Defn 10** (Probability Law). A *probability law* for a random experiment  $E$ , with sample space  $S$ , and an event class  $\mathcal{F}$  is a rule that assigns to each event  $A \in \mathcal{F}$  a number  $P[A]$ , called the probability of  $A$  that satisfies the axioms:

Axiom I:  $0 \leq P[A]$

Axiom II:  $P[S] = 1$

Axiom III: If  $A \cap B = \emptyset$ , then  $P[A \cup B] = P[A] + P[B]$

Axiom III': If  $A_1, A_2, \dots$  is a sequence of events such that  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ , then  $P[\bigcup_{k=1}^{\infty} A_k] = \sum_{k=1}^{\infty} P[A_k]$

### 3.2 Probability Law Corollaries

Axiom I:  $0 \leq P[A]$

Axiom II:  $P[S] = 1$

Axiom III: If  $A \cap B = \emptyset$ , then  $P[A \cup B] = P[A] + P[B]$

Axiom III': If  $A_1, A_2, \dots$  is a sequence of events such that  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ , then  $P[\bigcup_{k=1}^{\infty} A_k] = \sum_{k=1}^{\infty} P[A_k]$

**Corollary 3.1.**  $P[A^C] = 1 - P[A]$

**Corollary 3.2.**  $P[A] \leq 1$

**Corollary 3.3.**  $P[\emptyset] = 0$

**Corollary 3.4.** If  $A_1, A_2, \dots, A_n$  are pairwise mutually exclusive ( $A_1 \cap A_2 \cap \dots \cap A_n = \emptyset$ ), then  $P[\bigcup_{k=1}^n A_k] = \sum_{k=1}^n P[A_k]$  for  $n \geq 2$

**Corollary 3.5.**  $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

**Corollary 3.6.**  $P[A \cup B] = \sum_{j=1}^n P[A_j] - \sum_{j < k} P[A_j \cap A_k] + \dots + (-1)^{n+1} P[A_1 \cap \dots \cap A_n]$

**Corollary 3.7.** If  $A \subset B$ , then  $P[A] \leq P[B]$

### 3.3 Conditional Probability

**Defn 11** (Conditional Probability). The *conditional probability* of event  $A$  **GIVEN THAT** event  $B$  occurred is denoted  $P[A|B]$  and is defined as

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad (3.1)$$

**Theorem 1** (Theorem of Total Probability). Let  $B_1, B_2, \dots, B_n$  be mutually exclusive events whose union equals the sample space  $S$ , i.e.  $B_1, B_2, \dots, B_n$  is a partition of  $S$ .

**Defn 12** (Baye's Rule). Let  $B_1, B_2, \dots, B_n$  be a partition of sample space  $S$ .

$$P[B_j|A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A|B_j] * P[B_j]}{\sum_{k=1}^n P[A|B_k] * P[B_k]} \quad (3.2)$$

### 3.4 Event Independence

**Defn 13** (Independent). Two events  $A$  and  $B$  are *independent* if

$$P[A \cap B] = P[A] * P[B], P[A] \neq 0, P[B] \neq 0 \quad (3.3)$$

- If  $A \cap B = \emptyset$ , the  $A$  and  $B$  are **dependent**.
- If checking for independence between more than 2 events, you must check each pair, each triple, etc. until you check the independence of each event against each other. For 3 events,  $A, B, C$ :
  - Check  $P[A \cap B \cap C] = P[A] * P[B] * P[C]$
  - Also need to check:
    1.  $P[A \cap B] = P[A] * P[B]$
    2.  $P[B \cap C] = P[B] * P[C]$
    3.  $P[A \cap C] = P[A] * P[C]$

If 2 events  $A$  and  $B$  are independent, then their complements are also independent. This is shown in Event Independence.

*Independence of Complements of Events.* We assumed that  $A$  and  $B$  were independent, so  $P[A \cap B] = P[A] \cdot P[B]$ . There are 2 more facts we will need:

Fact 1:  $P[B] + P[B^C] = 1$

Fact 2:  $P[A \cap B^C] + P[A \cap B] = P[A]$

From Fact 1, we have:

$$P[A \cap B] = P[A] \cdot (1 - P[B^C])$$

From Fact 2, we have  $P[A \cap B] = P[A] - P[A \cap B^C]$ . Substituting these into the equation above:

$$\begin{aligned} P[A] - P[A \cap B^C] &= P[A] \cdot (1 - P[B^C]) \\ P[A] - P[A \cap B^C] &= P[A] - P[A] \cdot P[B^C] \\ -P[A \cap B^C] &= -P[A] \cdot P[B^C] \\ P[A \cap B^C] &= P[A] \cdot P[B^C] \end{aligned}$$

$\therefore A$  and  $B^C$  are independent, according to the definition of Independent events in Equation (3.3). ■

## 4 Counting

### 4.1 Sampling *with* Replacement *with* Order

**Defn 14.** Choose  $k$  elements in succession with replacement between selections, from a population of  $n$  distinct objects, where  $k$  needs to have no relation to  $n$ .

$$\frac{n}{\text{First}} * \frac{n}{\text{Second}} * \frac{n}{\text{Third}} * \dots * \frac{n}{\text{kth Item}} = n^k \quad (4.1)$$

### 4.2 Sampling *without* Replacement *with* Order

**Defn 15.** Choose  $k$  elements in succession without replacement from a population of  $n$  distinct objects, where  $k \leq n$

$$\frac{n}{\text{First}} * \frac{n-1}{\text{Second}} * \frac{n-2}{\text{Third}} * \dots * \frac{n-k+1}{\text{kth Item}} \quad (4.2)$$

#### 4.2.1 Permutations

**Defn 16** (Permutation). *Permutations* are special cases of Sampling *without* Replacement *with* Order, where  $k = n$

$$\frac{n}{\text{First}} * \frac{n-1}{\text{Second}} * \frac{n-2}{\text{Third}} * \dots * \frac{2}{\text{Second to last}} * \frac{1}{\text{last}} = n! \quad (4.3)$$

### 4.3 Sampling *with* Replacement *without* Order

**Defn 17.** Pick  $k$  objects from a set of  $n$  distinct object with replacement. Record the result without order. The total number of ways to do this is given in Equation (4.4).

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1} \quad (4.4)$$

### 4.4 Sampling *without* Replacement *without* Ordering

**Defn 18.** Pick  $k$  objects from a set of  $n$  distinct objects without replacement. Record the results with without order. We call the resulting subset of  $k$  selected objects a “combination of size  $k$ .” The number of ways to choose  $k$  items out of  $n$  items is given in Equation (4.5). Also said  $n$  choose  $k$ :

$$\binom{n}{k} = \frac{n * (n-1) * (n-2) * \dots * (n-k+1)}{k!} = \frac{n!}{k! (n-k)!} \quad (4.5)$$

$$\binom{n}{k} = \binom{n}{n-k} \quad (4.6)$$

## 5 Single Discrete Random Variables

**Defn 19** (Random Variable). A *random variable*  $X$  is a function that assigns a real number  $X(\zeta)$  to each outcome  $\zeta$  in the sample space of the random experiment.

**Defn 20** (Discrete Random Variable). A *discrete random variable* is a random variable that assumes values in a countable set. For example, the number of heads in 3 coin flips is a discrete random variable.

### 5.1 Probability Mass Function (PMF)

**Defn 21** (Probability Mass Function). The *probability mass function (PMF)* of a discrete random variable  $X$  is defined as:

$$p_X(x) = P[X = x] \quad (5.1)$$

Using the coin example from the definition of a Discrete Random Variable,

$$p_X(x) = \begin{cases} \frac{1}{8} & x = 0 \\ \frac{3}{8} & x = 1 \\ \frac{3}{8} & x = 2 \\ \frac{1}{8} & x = 3 \end{cases} \quad (5.2)$$

#### 5.1.1 Properties of Probability Mass Functions

(i) 
$$p_X(x) \geq 0, \forall x \in \mathbb{R} \quad (5.3)$$

(ii) 
$$\sum_{x \in S_X} p_X(x) = 1 \quad (5.4)$$

(iii) 
$$P[x \in B] = \sum_{x \in B} p_X(x), \text{ where } B \subset S_X \quad (5.5)$$

### 5.2 Expected Value/Mean of Single Discrete Random Variable

**Defn 22** (Expected Value/Mean of Single Discrete Random Variable). The *expected value* or *mean* of a single discrete random variable  $X$  is defined by

$$m_X = \mathbb{E}[X] = \sum_{x \in S_X} x \cdot p_X(x) \quad (5.6)$$

*Remark 22.1.* If  $X$  is countably infinite, you will have an infinite series that exists only if

$$\sum_{s \in S_X} |x| \cdot p_X(x) \quad (5.7)$$

is absolutely convergent.

#### 5.2.1 Properties of Expected Values

**Defn 23** (Linearity of Expectation). Let  $Y = X_1 + X_2$

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] \quad (5.8)$$

This can be generalized to

$$\mathbb{E}\left[\sum_{i=1}^k x_i\right] = \sum_{i=1}^k \mathbb{E}[X_i] \quad (5.9)$$

(i) 
$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] \quad (5.10)$$

(ii) 
$$\mathbb{E}[g(X)] = \sum_{s \in S_X} g(x) \cdot p_X[X] \quad (5.11)$$

(iii)

$$\mathbb{E}[cg(X)] = c\mathbb{E}[g(X)] \quad (5.12)$$

(iv)

$$\mathbb{E}[g_1(X) + g_2(X) + \dots + g_m(X)] = \sum_{i=1}^m \mathbb{E}[g_i(X)] \quad (5.13)$$

### 5.2.2 Moments of Random Variable

**Defn 24** (Moment). The *moment* of a random variable,  $X$  is defined as the expectation of the random variable raised to the moment.

$$\begin{aligned} \mathbb{E}[X^1] &= \text{First Moment} \\ \mathbb{E}[X^2] &= \text{Second Moment} \\ &\vdots \\ \mathbb{E}[X^k] &= \text{kth Moment} \end{aligned} \quad (5.14)$$

### 5.3 Variance of Single Discrete Random Variable

**Defn 25** (Variance). The *variance* of a single discrete random variable  $X$  is defined as:

$$\mathbb{E}[(X - \mathbb{E}[X])^2] \quad (5.15)$$

$$\text{VAR}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \quad (5.16)$$

and is denoted as  $\sigma_X^2$ , or as the operator  $\text{VAR}[X]$ .

*Remark 25.1.* If  $X$  is a random variable, and  $c$  is some constant coefficient, then:

$$\text{VAR}[cX] = c^2 \text{VAR}[X] \quad (5.17)$$

**Defn 26** (Standard Deviation). The standard deviation of a random variable  $X$  is:

$$\sigma_X = \sqrt{\text{VAR}[X]} \quad (5.18)$$

### 5.4 Conditional Probability Mass Function

**Defn 27** (Conditional Probability Mass of Function). Let  $X$  be a discrete random variable, with PMF  $p_X(x)$  and let  $C$  be the event with non-zero probability, i.e.  $P[C] > 0$ . The *conditional probability mass function of  $X$  given  $C$*  (*Conditional PMF*) is defined as:

$$p_{X|C}(x|C) = P[X = x | C] \text{ for } x \in \mathbb{R} \quad (5.19)$$

*Remark 27.1.* The conditional PMF,  $p_{X|C}(x|C)$ , satisfies **all** properties of Properties of Probability Density Functions.

### 5.5 Conditional Expected Value of Single Discrete Random Variable

**Defn 28** (Conditional Expected Value of Discrete Random Variable). The *conditional expected value of the discrete random variable  $X$  given  $B$*  is defined as:

$$m_{X|B} = \mathbb{E}[X | B] = \sum_{x \in S_X} x \cdot p_X(x | B) \quad (5.20)$$

### 5.6 Conditional Variance of Single Discrete Random Variable

**Defn 29** (Conditional Variance of Discrete Random Variable). The *conditional variance of a discrete random variable  $X$  given event  $B$*  as defined as:

$$\begin{aligned} \sigma_{X|B}^2 &= \text{VAR}[X | B] \\ &= \mathbb{E}[(X - \mathbb{E}[X | B])^2 | B] \\ &= \sum_{x \in S_X} (x - m_{X|B})^2 \cdot p_X(x | B) \\ \text{VAR}[X | B] &= \mathbb{E}[X^2 | B] - (\mathbb{E}[X | B])^2 \end{aligned} \quad (5.21)$$



## 6 Single Continuous Random Variables

**Defn 30** (Random Variable). Consider a random experiment with sample space  $S$  and event class  $\mathcal{F}$ . A *random variable*  $X$  is a function from the sample space  $S$  to the real line  $\mathbb{R}$  with the property the set  $A_b = \{\zeta : X|\zeta \leq b\}$  is in  $\mathcal{F}$  for every  $b$  in  $\mathbb{R}$ .

**Defn 31** (Continuous Random Variable). A *continuous random variable* is a random variable whose Cumulative Distribution Function (CDF) is continuous everywhere.

### 6.1 Cumulative Distribution Function (CDF)

**Defn 32** (Cumulative Distribution Function). *Cumulative Distribution Function (CDF)* of a random variable  $X$  is defined as the probability of the event  $\{X \leq x\}$ .

$$F_X(x) = P[X \leq x] \text{ for } -\infty < x < \infty \quad (6.1)$$

#### 6.1.1 Properties of Cumulative Distribution Functions

(i) 
$$x \leq F_X(x) \leq 1 \quad (6.2)$$

(ii) If you include the whole sample space, you should end up with 1.

$$\lim_{x \rightarrow \infty} F_X(x) = 1 \quad (6.3)$$

(iii) If you exclude the whole sample space, you should end up with 0.

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad (6.4)$$

(iv)  $F_X(x)$  is non-decreasing.

$$F_X(a) \leq F_X(b) \text{ if } a \leq b \quad (6.5)$$

(v) The CDF is continuous from the right.

$$F_b = \lim_{h \rightarrow 0} F_X(b+h) \text{ where } h > 0 \quad (6.6)$$

(vi) 
$$P[a < X \leq b] = F_X(b) - F_X(a) \quad (6.7)$$

(vii) The probability at a point in a CDF. (This usually ends up being 0).

$$P[X = b] = F_X(b) - F_X(b^-) \quad (6.8)$$

(viii) The probability of the event **not** occurring.

$$P[X > x] = 1 - P[X \leq x] = 1 - F_X(x) \quad (6.9)$$

#### 6.1.2 Conditional Cumulative Distribution Function

**Defn 33** (Conditional Cumulative Distribution Function). The *conditional cumulative distribution function (Conditional CDF)* of  $X$  given  $C$  is defined by:

$$F_{X|C}(x|C) = \frac{P[\{X = x\} | C]}{P[C]} \quad (6.10)$$

*Remark 33.1.* The conditional CDF,  $F_{X|C}(x|C)$  satisfies **all** Properties of Cumulative Distribution Functions.

### 6.2 Probability Density Function (PDF)

**Defn 34** (Probability Density Function). The *probability density function (PDF)* of a random variable  $X$ , if it exists, is defined as the derivative of the CDF of  $X$ .

$$f_X(x) = \frac{d}{dx} F_X(x) \quad (6.11)$$

*Remark 34.1.* Both discrete and continuous random variables can have PDFs, however, the discrete random variable will have a discontinuous PDF.

*Remark 34.2.* It is possible to construct a random variable that has a Cumulative Distribution Function (CDF), but an undefined Probability Density Function (PDF).

*Remark 34.3.* This is an alternate, more useful way to specify the probability law described by the Cumulative Distribution Function (CDF).

### 6.2.1 Properties of Probability Density Functions

These properties apply to PDFs of continuous random variables, and may not hold true for other types of random variables.

- (i) The associated CDF is non-decreasing, a Properties of Cumulative Distribution Functions.

$$f_X(x) \geq 0 \quad (6.12)$$

- (ii) Since the definition of the PDF is that it's the derivative of the CDF, integrating the space over the PDF will yield the CDF.

$$P[a \leq X \leq b] = \int_a^b f_X(x) dx = F_X(b) - F_X(a) \quad (6.13)$$

- (iii) The value of a location in CDF is the integral of the PDF over the area.

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad (6.14)$$

- (iv) Including the whole sample space should yield 1.

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (6.15)$$

*Remark.* Any non-negative, piecewise continuous function  $g(x)$  with finite  $\int_{-\infty}^{\infty} g(x) dx = C$  can be used to form a PDF.

### 6.2.2 Conditional Probability Density Function

**Defn 35** (Conditional Probability Density Function). The *conditional probability density function (Conditional PDF)* of  $X$  given  $C$  is defined by:

$$f_{X|C}(x|C) = \frac{d}{dx} F_{X|C}(x|C) \quad (6.16)$$

*Remark 35.1.* The conditional PDF,  $f_{X|C}(x|C)$  satisfies **all** Properties of Probability Density Functions.

## 6.3 Expected Value of Single Continuous Random Variable

**Defn 36** (Expected Value/Mean of Random Variable). The *expected value of a random variable  $X$* , denoted  $\mathbb{E}[X]$  is defined as:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t f_X(t) dt \quad (6.17)$$

*Remark 36.1.* This works with **all** random variables, or general random variables.

*Remark 36.2.*  $\mathbb{E}[X]$  is defined if the integral in Equation (6.17) converges absolutely. This means:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t f_X(t) dt < \infty$$

### 6.3.1 Properties of Expected Value

- (i) The expected value of a function of a random variable.

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(t) \cdot f_X(t) dt \quad (6.18)$$

- (ii) Expectation of a constant,  $c$ , should be the constant itself.

$$\mathbb{E}[c] = c \quad (6.19)$$

- (iii) Sum of a random variable,  $X$ , and a constant,  $c$ , is the same as the sum of the expectation of the random variable and the constant.

$$\mathbb{E}[X + c] = \mathbb{E}[X] + \mathbb{E}[c] \quad (6.20)$$

- (iv) Linearity of Expectations for random variables

$$\mathbb{E}[a_0 + a_1 X + a_2 X^2 + \dots + a_n X^n] = a_0 + a_1 \mathbb{E}[X] + a_2 \mathbb{E}[X^2] + \dots + a_n \mathbb{E}[X^n] \quad (6.21)$$

## 6.4 Variance of Single Continuous Random Variable

**Defn 37** (Variance of Random Variable). The *variance* of the random variable  $X$  is defined by:

$$\sigma^2 = \text{VAR}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] \quad (6.22)$$

*Remark 37.1.* This holds true for **all** types of random variables; discrete, continuous, and mixed.

**Defn 38** (Standard Deviation). The *standard deviation* of a random variable  $X$ , denoted by:

$$\sigma = \text{STD}[X] = \sqrt{\text{VAR}[X]} \quad (6.23)$$

*Remark 38.1.* This holds true for **all** types of random variables; discrete, continuous, and mixed.

## 6.5 Gaussian/Normal Random Variable

**Defn 39** (Gaussian/Normal Random Variable). The *Gaussian or normal random variable* is the classic “bell curve” probability distribution. It is usually described as  $X \sim N(\mu, \sigma^2)$ .  $\mu$  is  $\mathbb{E}[X]$  and  $\sigma^2$  is how narrow/sharp the bell is. A Gaussian Random Variable has a PDF of:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R} \quad (6.24)$$

**Defn 40** (Standard Normal Distribution). The *standard normal distribution* is just a specific Gaussian/Normal Random Variable. The standard normal distribution is a Gaussian/Normal Random Variable with  $\mu = 0, \sigma^2 = 1$ .

*Remark 40.1.* The CDF of the Standard Normal Distribution is denoted with  $\Phi$ .

To find the probability of something for a Gaussian Random Variable, you would end up converting it to the Standard Normal Distribution. If  $X \sim N(\mu, \sigma^2)$  and  $Y \sim N(0, 1)$ ,

$$\begin{aligned} P[a \leq x \leq b] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{1}{2}y^2} dy \\ &= P\left[\frac{a-\mu}{\sigma} \leq Y \leq \frac{b-\mu}{\sigma}\right] \\ &= F_Y\left(\frac{b-\mu}{\sigma}\right) - F_Y\left(\frac{a-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned} \quad (6.25)$$

### 6.5.1 Q-Function

**Defn 41** (Q-Function). The *Q-Function* is primarily used in electrical engineering. It is defined as:

$$\begin{aligned} Q &= 1 - \Phi(x) \\ &= \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt \end{aligned} \quad (6.26)$$

*Remark 41.1.*

$$Q(Z) = 1 - f_Z(z) \quad (6.27)$$

## 6.6 Markov Inequality

**Defn 42** (Markov Inequality). Let  $X$  be a non-negative random variable with  $\mathbb{E}[X] < \infty$ . The *Markov Inequality* states that:

$$P[X \geq a] \leq \frac{\mathbb{E}[X]}{a} \quad (6.28)$$

*Proving the Markov Inequality.*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

Because we defined  $X \geq 0$ , we change the lower bound to 0.

$$\mathbb{E}[X] = \int_0^{\infty} x f_X(x) dx$$

We then split the integral up around some point,  $a$ .

$$\mathbb{E}[X] = \int_0^a x f_X(x) dx + \int_a^{\infty} x f_X(x) dx$$

Since the first integral is integrating over a non-negative function, the integral is also non-negative.

$$\int_0^a x f_X(x) dx + \int_a^{\infty} x f_X(x) dx \geq \int_a^{\infty} x f_X(x) dx$$

$$\mathbb{E}[X] \geq \int_a^{\infty} x f_X(x) dx$$

Because  $x > a$ , we can pull a term out of  $f_X(x)$

$$\mathbb{E}[X] \geq \int_a^{\infty} a f_X(x) dx$$

Because  $a$  is a constant, we pull it out of the integral,

$$\mathbb{E}[X] \geq a \int_a^{\infty} f_X(x) dx$$

Then, we end up with an integral that is the definition of the probability of a continuous random variable.

$$\mathbb{E}[X] \geq a P[X \geq a]$$

$$\therefore \mathbb{E}[X] \geq a P[X \geq a]$$

■

## 6.7 Chebychev Inequality

**Defn 43** (Chebychev Inequality). Let  $X$  be a non-negative random variable with  $\mathbb{E}[X] < \infty$ . The *Chebychev Inequality* states that:

$$P[|X - \mu| \geq a] \leq \frac{\sigma^2}{a^2} \quad (6.29)$$

*Proving the Chebychev Inequality.*

$$P[(X - \mu)^2 \geq a^2] \leq \frac{\mathbb{E}[(X - \mu)^2]}{a^2}$$

Because  $X - \mu = \sigma$ , we replace it.

$$P[(X - \mu)^2 \geq a^2] \leq \frac{\mathbb{E}[\sigma^2]}{a^2}$$

■

## 7 Multiple Random Variables

### 7.1 Joint Probability Mass Function

**Defn 44** (Joint Probability Mass Function). The *joint probability mass function (joint PMF)* of 2 discrete random variables  $X, Y$  is defined as:

$$p_{X,Y} = P[\{X = x\} \cap \{Y = y\}] \text{ for all } x, y \in S_{X,Y} \quad (7.1)$$

- This satisfies ALL propoerties of single random variable PMFs

### 7.1.1 Marginal Probability Mass Function

**Defn 45** (Marginal Probability Mass Function). Given a joint PMF of discrete random variables  $X, Y$ , the *Marginal Probability Mass Function (Marginal PMF)* of  $X$  is defined as:

$$p_X(x_i) = P[X = x_i] \text{ for } x_i \in S_X \quad (7.2)$$

and is calculated as:

$$p(x_i) = \sum_{y \in S_Y} p_{X,Y}(x_i, y) \quad (7.3)$$

## 7.2 Joint Cumulative Distribution Function

**Defn 46** (Joint Cumulative Distribution Function). The *Joint Cumulative Distribution Function (Joint CDF)* of  $X$  and  $Y$  is defined as the probability of the event  $\{X \leq x\} \cap \{Y \leq y\}$

$$\begin{aligned} F_{X,Y}(x, y) &= P[\{X \leq x\} \cap \{Y \leq y\}] \text{ for all } (x, y) \in \mathbb{R}^2 \\ &= P[\{X \leq x\}, \{Y \leq y\}] \end{aligned} \quad (7.4)$$

### 7.2.1 Properties of Joint Cumulative Distribution Functions

(i)  $F_{X,Y}(x, y)$  is non decreasing.

$$F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2) \text{ if } x_1 \leq x_2 \text{ and } y_1 \leq y_2 \quad (7.5)$$

(ii)

$$\begin{aligned} \lim_{y \rightarrow -\infty} F_{X,Y}(x, y) &= 0 \\ \lim_{x \rightarrow -\infty} F_{X,Y}(x, y) &= 0 \\ \lim_{(x,y) \rightarrow (\infty, \infty)} F_{X,Y}(x, y) &= 1 \end{aligned} \quad (7.6)$$

(iii) The Marginal CDFs can be obtained from the Joint CDF by removing restrictions for all but one variable.

$$\begin{aligned} F_X(x) &= P[\{X \leq x\}, \{Y \text{ is anything}\}] \\ &= P[\{X \leq x\}, \{-\infty \leq y \leq \infty\}] \\ &= \lim_{y \rightarrow \infty} F_{X,Y}(x, y) \\ F_Y(y) &= \lim_{x \rightarrow \infty} F_{X,Y}(x, y) \end{aligned} \quad (7.7)$$

(iv) The Joint CDF is continuous from  $\infty$  to  $-\infty$ .

$$\begin{aligned} \lim_{x \rightarrow a^+} F_{X,Y}(x, y) &= F_{X,Y}(a, y) \\ \lim_{y \rightarrow b^+} F_{X,Y}(x, y) &= F_{X,Y}(x, b) \end{aligned} \quad (7.8)$$

(v) The probability of the “rectangle”  $\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\}$

$$\begin{aligned} P[\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\}] &= P[\{X \leq x_2, Y \leq y_2\}] - P[\{X \leq x_1, Y \leq y_2\}] - \\ &\quad P[\{X \leq x_2, Y \leq y_1\}] + P[\{X \leq x_1, Y \leq y_1\}] \\ &= F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1) \end{aligned} \quad (7.9)$$

### 7.2.2 Marginal Cumulative Distribution Function

**Defn 47** (Marginal Cumulative Distribution Function). We obtain the *Marginal Cumulative Distribution Functions (Marginal CDFs)* by removing the constraint on one of the variables.

$$\begin{aligned} F_X(x) &= P[\{X \leq x\}, \{Y \text{ is anything}\}] \\ &= P[\{X \leq x\}, \{-\infty \leq y \leq \infty\}] \\ &= \lim_{y \rightarrow \infty} F_{X,Y}(x, y) \\ F_Y(y) &= \lim_{x \rightarrow \infty} F_{X,Y}(x, y) \end{aligned} \quad (7.10)$$

### 7.3 Joint Probability Density Function

**Defn 48** (Joint Probability Density Function). We say that  $X, Y$  are jointly continuous if the probabilities of events involving  $X$  and  $Y$  can be expressed as an integral of a *Joint Probability Density Function (Joint PDF)*.

i.e. There exists some nonnegative function  $f_{X,Y}(x, y)$ , which we call the joint PDF, that is defined on the real plane such that for every event  $B$  which is a subset of the  $xy$  plane

$$P[(X, Y) \text{ in } B] = \iint_B f_{X,Y}(x, y) dx dy \quad (7.11)$$

*Remark 48.1.* The probability mass of an event is found by integrating the PDF over the region in the  $xy$  plane corresponding to your event.

#### 7.3.1 Properties of Joint Probability Density Functions

$$\iint_B f_{X,Y}(x, y) = 1 \quad (7.12)$$

$$x \geq 0, y \geq 0 \forall x \forall y \quad (7.13)$$

$$(7.14)$$

#### 7.3.2 Facts about Joint PDFs

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) = 1 \quad (7.15)$$

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) dt ds \quad (7.16)$$

$$f_{X,Y} = \frac{\partial^2 f_{X,Y}(x, y)}{\partial x \partial y} \quad (7.17)$$

$$(7.18)$$

#### 7.3.3 Marginal PDF

**Defn 49** (Marginal Probability Density Function). The *Marginal Probability Density Functions (Marginal PDFs)*  $f_X(x)$  and  $f_Y(y)$  are obtained by taking the derivative of the marginal CDFs.

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx} \int_{-\infty}^x \left[ \int_{-\infty}^{\infty} f_{X,Y}(s, t) dt ds \right] \\ &= \frac{d}{dx} \int_{-\infty}^x \int_{-\infty}^{\infty} f_{X,Y}(s, t) dt ds \end{aligned} \quad (7.19)$$

Simplified with Second Fundamental Theorem of Calculus

$$\begin{aligned} &= \int_{-\infty}^{\infty} f_{X,Y}(x, t) dt \\ f_X &= \int_{-\infty}^{\infty} f_{X,Y}(x, t) dt \end{aligned}$$

### 7.4 Independence of Multiple Random Variables

**Defn 50** (Independent Random Variables).  $X$  and  $Y$  are independent random variables if **ANY** event  $A_1$  defined in terms of  $S$  is independent of **ANY** event  $A_2$  defined in terms of  $Y$ .

$$P[X \in A_1, Y \in A_2] = P[X \in A_1] * P[Y \in A_2] \quad (7.20)$$

There are 3 ways to phrase this:

1. For discrete random variables  $X$  and  $Y$ ,  $X$  and  $Y$  are independent if and only if:

$$p_{X,Y}(x, y) = p_X(x) * p_Y(y) \quad (7.21)$$

2. For general random variables  $X$  and  $Y$ ,  $X$  and  $Y$  are independent if and only if:

$$F_{X,Y}(x, y) = F_X(x) * F_Y(y) \quad (7.22)$$

3. For (continuous) random variables  $X$  and  $Y$ ,  $X$  and  $Y$  are independent if and only if:

$$f_{X,Y}(x, y) = f_X(x) * f_Y(y) \quad (7.23)$$

You can prove Independence of Multiple Random Variables, Equation (7.21).

*Independence of Discrete Random Variables with PMF.* ■

**Theorem 2** (Independence of Random Functions). *If random variables  $X$ ,  $Y$  are independent, then  $g(X)$  and  $h(Y)$  are also independent.*

## 7.5 Expected Value of Functions with 2 Random Variables

**Defn 51** (Expectation of a Function with 2 Random Variables). Let  $Z$  be a random variable described by the function  $Z = g(X, Y)$ .

$$\mathbb{E} = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y) dx dy & \text{if } X \text{ and } Y \text{ are jointly continuous} \\ \sum_{i \in S_X} \sum_{j \in S_Y} g(x_i, y_j) \cdot p_{X,Y}(x_i, y_j) & \text{if } X \text{ and } Y \text{ are both discrete} \end{cases} \quad (7.24)$$

*Remark 51.1* (Expected Value of Sum of Random Variables). You **do not** need to assume independence to say:

$$\mathbb{E}[X_1 + X_2 + \dots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n] \quad (7.25)$$

*Remark 51.2* (Expected Value of Product of Random Variables). If  $X$  and  $Y$  are independent, then

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)] \quad (7.26)$$

## 7.6 Joint Moments, Correlation, and Covariance

### 7.6.1 Joint Moments

**Defn 52** (The  $j, k$ th Moment). The  $j, k$ th moment of  $X$  and  $Y$  is:

$$\mathbb{E}[X^j Y^k] = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k \cdot f_{X,Y}(x, y) dx dy & \text{if } X, Y \text{ are jointly continuous} \\ \sum_{i \in S_X} \sum_{\ell \in S_Y} x_i^j y_\ell^k \cdot p_{X,Y}(x_i, y_\ell) & \text{if } X, Y \text{ are discrete} \end{cases} \quad (7.27)$$

### 7.6.2 Covariance

**Defn 53** (Covariance). The *covariance* of  $X$  and  $Y$  is denoted:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \quad (7.28)$$

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (7.29)$$

### 7.6.3 Correlation

**Defn 54** (Correlation). The *Correlation* of  $X$  and  $Y$  is defined as the 1, 1 moment, i.e.  $\mathbb{E}[X^1 Y^1]$ .

*Remark 54.1.* If  $X, Y$  are such that  $\mathbb{E}[X^1 Y^1] = 0$ , then we say that  $X, Y$  are *orthogonal*.

*Remark 54.2.* If  $X, Y$  are such that  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ , then  $X$  and  $Y$  are *uncorrelated*.

**Defn 55** (Correlation Coefficient). The *correlation coefficient* of  $X, Y$  is defined as

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} \quad (7.30)$$

*Remark 55.1.*  $\rho_{X,Y}$  only ranges  $-1 \leq \rho_{X,Y} \leq 1$

*Remark 55.2.* The closer  $\rho_{X,Y}$  is to +1, the closer  $X$  and  $Y$  are to having a positive linear relationship (Positive slope).

The closer  $\rho_{X,Y}$  is to -1, the closer  $X$  and  $Y$  are to having a negative linear relationship (Negative slope).

If  $\rho_{X,Y} = 0$ , the  $\text{Cov}[X, Y] = 0$ , which means that  $X$  and  $Y$  are *uncorrelated*.

*Remark 55.3.* If  $X, Y$  are independent, then they are uncorrelated; but if  $X$  and  $Y$  are uncorrelated, **they are not always independent**.

## 7.7 Conditional Probability Functions

There are 3 major cases for these:

1. 2 Discrete Random Variables
2. 1 Discrete and 1 Continuous Random Variable
3. 2 Continuous Random Variables

### 7.7.1 2 Discrete Random Variables

**Defn 56** (Conditional Probability Mass Function). The *conditional Probability Mass Function (Conditional PMF)* of  $Y$  given that  $X = x$  is:

$$p_Y(y|x) = \frac{P[\{Y = y\} \cap \{X = x\}]}{P[X = x]} = \frac{p_{X,Y}(x, y)}{p_X(x)} \quad (7.31)$$

*Remark 56.1.* This also implies that

$$p_{X,Y}(x, y) = p_Y(y|x) \cdot p_X(x) \quad (7.32)$$

*Remark 56.2.* If  $X$  and  $Y$  are *independent*, then:

$$p_X(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)} = \frac{p_X(x) p_Y(y)}{p_X(x)} = p_Y(y) \quad (7.33)$$

*Remark 56.3.* The Conditional Probability Mass Function of 2 discrete random variables satisfies all Properties of Probability Mass Functions.

### 7.7.2 1 Discrete and 1 Continuous Random Variable

For this section, let  $X$  be a discrete random variable and  $Y$  a continuous random variable.

**Defn 57** (Conditional Cumulative Distribution Function). The *conditional Cumulative Distribution Function (Conditional CDF)* of  $Y$  given that  $X = x$  is:

$$F_Y(y|x) = P[Y \leq y | X = x] = \frac{P[\{Y \leq y\} \cap \{X = x\}]}{P[X = x]} \quad (7.34)$$

*Remark 57.1.* If  $X$  and  $Y$  are *independent*, then:

$$F_Y(y|x) = \frac{F_{X,Y}(x, y)}{p_X(x)} = \frac{F_Y(y) p_X(x)}{p_X(x)} = F_Y(y) \quad (7.35)$$

This also means that:

$$P[Y \leq y | X = x] = P[Y \leq y] \cdot P[X = x]$$

*Remark 57.2.* The similar relations for independent random variables with their conditional and marginal probability functions does not hold true with this.

*Remark 57.3.* The Conditional Cumulative Distribution Function of 1 discrete random variable and 1 continuous random variable satisfies all Properties of Cumulative Distribution Functions.

**Defn 58** (Conditional Probability Density Function). The *conditional Probability Distribution Function (Conditional PDF)* of  $Y$  given  $X = x$  is

$$f_Y(y|x) = \frac{d}{dy} F_Y(y|x) \quad (7.36)$$

This also means,

$$P[Y \leq y | X = x] = \int_{y \in A} f_Y(y|x) dy$$

*Remark 58.1.* The Conditional Probability Density Function of 1 discrete random variable and 1 continuous random variable satisfies all Properties of Probability Density Functions.



### 7.7.3 2 Continuous Random Variables

**Defn 59** (Conditional Cumulative Distribution Function). The *conditional Cumulative Distribution Function (Conditional CDF)* of  $Y$  given  $X = x$  for  $X$  and  $Y$  continuous random variables is:

$$F_Y(y|x) = \lim_{h \rightarrow 0} F_Y(y|x < X \leq (x+h)) = \frac{\int_{-\infty}^y f_{X,Y}(x,v) dv}{f_X(x)} \quad (7.37)$$

*Remark 59.1.* The Conditional Cumulative Distribution Function of 2 continuous random variables satisfies all Properties of Cumulative Distribution Functions.

*Remark 59.2.* The similar relations for the conditional and marginal probability functions do not hold up for 2 continuous random variables too well.

**Defn 60** (Conditional Probability Density Function). The *conditional Probability Density Function (Conditional PDF)* of  $Y$  given  $X = x$  for  $X$  and  $Y$  continuous random variables is:

$$f_Y(y|x) = \frac{d}{dy} F_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \quad (7.38)$$

*Remark 60.1.* If  $X$  and  $Y$  are independent, then:

$$f_X(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y) \quad (7.39)$$

*Remark 60.2.* The Conditional Probability Density Function of 2 continuous random variables satisfies all Properties of Probability Density Functions.

## 7.8 Conditional Expectation of Multiple Random Variables

**Defn 61** (Conditional Expectation). The *conditional expectation* of  $Y$  given  $X$  is:

$$\mathbb{E}[Y|X=x] = \int_{-\infty}^{\infty} y \cdot f_Y(y|x) dy \quad (7.40)$$

*Remark 61.1* (Special Case). There is a special case when **both**  $X$  and  $Y$  are discrete random variables.

$$\mathbb{E}[Y|X=x] = \sum_{y \in S_Y} y \cdot p_Y(y|x) \quad (7.41)$$

*Remark 61.2.* When calculating the Conditional Expectation of Multiple Random Variables, and they as for  $\mathbb{E}[Y|X=x]$ , that means you **must** consider all possible values that  $X$  can take. This can be generalized to the equation below.

$$\mathbb{E}[Y|X=x] = \sum_{x \in S_X} \left( \sum_{y \in S_Y} y \cdot p_Y(y|x) \right) \quad (7.42)$$

This can be described. You must take a single value for  $x$ , and take it over all  $y$ 's, then take the next value for  $x$ , until you have exhausted all values in both  $S_X$  and  $S_Y$ .

This can also be translated into the continuous case, but the discrete case is a little simpler to understand this generality.

*Remark 61.3.*  $\mathbb{E}[Y|X=x]$  is a function of  $X$ , so it can be written as  $g(x) = \mathbb{E}[Y|X=x]$ . Thus, we can also say

$$\mathbb{E}[g(X)] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y] \quad (7.43)$$

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \int_{-\infty}^{\infty} \mathbb{E}[Y|x] f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_Y(y|x) dy f_X(x) dx \quad (7.44a)$$

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \sum_{x \in S_X} \mathbb{E}[Y|x] p_X(x) = \sum_{x_j \in S_X} \sum_{y_i \in S_Y} y_i p_Y(y_i|x_k) p_X(x_j) \quad (7.44b)$$

*Prove Expectation of Conditional Expected Value.* ■

## 8 Random Vectors

Random Vectors are usually denoted:

$$\vec{X} = \langle X_1, X_2, X_3, \dots, X_n \rangle \quad (8.1)$$

## 8.1 Joint CDF of a Random Vector

$$\begin{aligned} F_{\vec{X}}(\vec{x}) &= F_{X_1, X_2, X_3, \dots, X_n}(x_1, x_2, x_3, \dots, x_n) \\ &= P[X_1 \leq x_1, X_2 \leq x_2, X_3 \leq x_3, \dots, X_n \leq x_n] \end{aligned} \quad (8.2)$$

## 8.2 Joint PDF of a Random Vector

$$f_{\vec{X}}(\vec{x}) = \frac{\partial^n F_{\vec{X}}(\vec{x})}{\partial x_1 \partial x_2 \partial x_3 \cdots \partial x_n} \quad (8.3)$$

### 8.2.1 Marginal PDF of a Random Vector

Integrate out the terms that you're not interested in.

$$f_{\vec{X}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\vec{X}}(\vec{x}) \partial x_2 \partial x_3 \cdots \partial x_n \quad (8.4)$$

For instance, say we want the marginal PDF of some function with respect to  $X_1$ ,  $X_3$ , and  $X_4$ .

$$f_{X_1, X_3, X_4}(x_1, x_3, x_4) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\vec{X}}(\vec{x}) \partial x_2 \partial x_5 \partial x_6 \cdots \partial x_n \quad (8.5)$$

## 8.3 Conditional Probability Functions of Random Vectors

This section is just an extension of Section 7.7, Conditional Probability Functions. There are 3 major cases for these:

1. Discrete Random Vectors
2. Mixed Random Vectors
3. Continuous Random Vectors

*Remark.* For the sections below, let  $\vec{Y} = \langle Y_1, Y_2, Y_3 \rangle$  and  $\vec{y} = \langle y_1, y_2, y_3 \rangle$ .

While I am using  $\vec{Y}$  and  $\vec{y}$ , these equations can be further generalized to higher dimensions. All that would be required for this is to keep track of everything.

### 8.3.1 Discrete Random Vectors

**Defn 62** (Conditional Probability Mass Function). The *conditional Probability Mass Function (Conditional PMF)* of  $Y_3$  given that  $Y = y$  is:

$$p_{Y_3}(y_3 | y_1, y_2) = \frac{P[\{Y_3 = y_3\} \cap (\{Y_1 = y_1\} \cap \{Y_2 = y_2\})]}{P[\{Y_1 = y_1\} \cap \{Y_2 = y_2\}]} = \frac{p_{\vec{Y}}(\vec{y})}{p_{Y_1, Y_2}(y_1, y_2)} \quad (8.6)$$

*Remark 62.1.* This also implies that

$$p_{\vec{Y}}(\vec{y}) = p_{Y_3}(y_3 | y_1, y_2) \cdot p_{Y_2}(y_2 | y_1) \cdot p_{Y_1}(y_1) \quad (8.7)$$

*Remark 62.2.* If all elements of  $\vec{Y}$  are *independent* (Remember that you need to check each subgroup too, like shown in Section 3.4), then:

$$p_{Y_3}(y_3 | y_1, y_2) = \frac{p_{\vec{Y}}(\vec{y})}{p_{Y_1, Y_2}(y_1, y_2)} = \frac{p_{Y_1, Y_2}(y_1, y_2) p_{Y_3}(y_3)}{p_{Y_1, Y_2}(y_1, y_2)} = p_{Y_3}(y_3) \quad (8.8)$$

*Remark 62.3.* The Conditional Probability Mass Function of 2 discrete random variables satisfies all Properties of Probability Mass Functions.

### 8.3.2 Mixed Random Vectors

### 8.3.3 Continuous Random Vectors

## 9 Statistics

### 9.1 Sums of Random Variables

**Defn 63** (Sum of Random Variables). The definition of a *sum of random variables* is given in Equation (9.1) below. Where  $X_i$  is a random variable,

$$S_n = \sum_{i=1}^n X_i = X_1 + X_2 + \cdots + X_n \quad (9.1)$$

### 9.1.1 Means and Variances of Sums of Random Variables

**Defn 64** (Mean of Sums of Random Variables). The *mean of sums of random variables* is the same as the *expected value of sums of random variables*.

$$\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i] \quad (9.2)$$

*Remark 64.1.* All the properties of Properties of Expected Values and/or Properties of Expected Value hold true here as well..

**Defn 65** (Variance of Sums of Random Variables). The definition of the *variance of sums of random variables* is the same as we have been using them previously, Variance of Single Discrete Random Variable and Variance of Single Continuous Random Variable.

$$\text{VAR}[S_n] = \text{VAR}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{VAR}[X_i] + \sum_{j=1}^n \sum_{\substack{k=1 \\ j \neq k}}^n \text{Cov}[X_j, X_k] \quad (9.3)$$

*Remark 65.1.* If  $X_1, X_2, \dots, X_n$  are independent, then:

$$\text{VAR}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{VAR}[X_i] \quad (9.4)$$

## A Reference Material

### A.1 Trigonometry

#### A.1.1 Trigonometric Formulas

$$\sin(\alpha) + \sin(\beta) = 2 \sin\left(\frac{\alpha + \beta}{2}\right) \cos\left(\frac{\alpha - \beta}{2}\right) \quad (\text{A.1})$$

$$\cos(\theta) \sin(\theta) = \frac{1}{2} \sin(2\theta) \quad (\text{A.2})$$

#### A.1.2 Euler Equivalents of Trigonometric Functions

$$\sin(x) = \frac{e^{ix} + e^{-ix}}{2i} \quad (\text{A.3})$$

$$\cos(x) = \frac{e^{ix} - e^{-ix}}{2i} \quad (\text{A.4})$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \quad (\text{A.5})$$

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \quad (\text{A.6})$$

### A.2 Calculus

#### A.2.1 Fundamental Theorems of Calculus

**Defn 66** (First Fundamental Theorem of Calculus). The *first fundamental theorem of calculus* states that, if  $f$  is continuous on the closed interval  $[a, b]$  and  $F$  is the indefinite integral of  $f$  on  $[a, b]$ , then

$$\int_a^b f(x) dx = F(b) - F(a) \quad (\text{A.7})$$

**Defn 67** (Second Fundamental Theorem of Calculus). The *second fundamental theorem of calculus* holds for  $f$  a continuous function on an open interval  $I$  and  $a$  any point in  $I$ , and states that if  $F$  is defined by

$$F(x) = \int_a^x f(t) dt,$$

then

$$\begin{aligned} \frac{d}{dx} \int_a^x f(t) dt &= f(x) \\ F'(x) &= f(x) \end{aligned} \quad (\text{A.8})$$