Contents

1 Probability Models

1.1 Relative Frequency

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

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

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

1.1.1 Properties of Relative Frequencies

(i)

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

(ii)

$$0 \leq N_k(n) \leq n \tag{1.3}$$

(iii)

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

(iv)

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

(v)

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

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

$$F_C = F_A(n) + F_B(n) \tag{1.7}$$

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).

$$\lim_{n\to\infty} f_k(n) = p_k \tag{1.8}$$

- 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, $x \in A \to x \in B$
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: $A = B \iff A \subset B$ 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 fo 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* $\bigcup_{k=1}^{n} A_k = A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n$ 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* $\bigcap_{k=1}^{n} A_k = A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n$ 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, … is a sequence of events such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$

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, … is a sequence of events such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$

Corollary 3.1. $P\left[A^C\right] = 1 - P[A]$

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

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

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

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

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

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

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \tag{3.1}$$

Theorem 3.1 (Theorem of Total Probability)**.** *Let B_1, B_2, …, B_n be mutually exclusive events whose union equals the sample space S, i.e. B_1, B_2, …, B_n is a* partition of S.

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

$$P[B_j|A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A|B_j] * P[B_j]}{\sum\limits_{k=1}^{n} P[A|B_k] * P[B_k]} \tag{3.2}$$

3.4 Event Independence

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

$$P[A \cap B] = P[A] * P[B], P[A] \neq 0, P[B] \neq 0 \tag{3.3}$$

- 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: $P[B] + P\left[B^C\right] = 1$

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

From Fact 1, we have:

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

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

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

$\therefore$ 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.

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

4.2 Sampling *without* Replacement *with* Order

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

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

4.2.1 Permutations

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

$$\frac{n}{First} * \frac{n-1}{Second} * \frac{n-2}{Third} * \dots * \frac{2}{-} * \frac{1}{-} * \dots * \frac{n-k-1}{kth \text{ Item}} = n! \tag{4.3}$$

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).

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1} \tag{4.4}$$

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:

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

$$\binom{n}{k} = \binom{n}{n-k} \tag{4.6}$$

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 ζ 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:

$$p_X(x) = P[X = x] \tag{5.1}$$

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

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

5.1.1 Properties of Probability Mass Functions

(i)

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

(ii)

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

(iii)

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

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

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

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

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

is absolutely convergent.

5.2.1 Properties of Expected Values

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

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] \tag{5.8}$$

This can be generalized to

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

(i)

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] \tag{5.10}$$

(ii)

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

(iii)

$$\mathbb{E}\left[cg\left(X\right)\right] = c\,\mathbb{E}\left[g\left(X\right)\right] \tag{5.12}$$

(iv)

$$\mathbb{E}\left[g_1\left(X\right) + g_2\left(X\right) + \ldots + g_m\left(X\right)\right] = \sum_{i=1}^{m} \mathbb{E}\left[g_i\left(X\right)\right] \tag{5.13}$$

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.

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

5.3 Variance of Single Discrete Random Variable

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

$$\mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right] \tag{5.15}$$

$$\mathrm{VAR}\left[X\right] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2 \tag{5.16}$$

and is denoted as σ_X^2, or as the operator $\mathrm{VAR}\left[X\right]$.

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

$$\mathrm{VAR}\left[cX\right] = c^2\,\mathrm{VAR}\left[X\right] \tag{5.17}$$

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

$$\sigma_X = \sqrt{\mathrm{VAR}\left[X\right]} \tag{5.18}$$

5.4 Conditional Probability Mass Function

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

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

Remark 27.1. The conditional PMF, $p_{X\,|\,C}\left(x\,|\,C\right)$, 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:

$$m_{X\,|\,B} = \mathbb{E}\left[X\,|\,B\right] = \sum_{x \in S_X} s \cdot p_X\left(x\,|\,B\right) \tag{5.20}$$

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:

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

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 $\{X \leq x\}$.

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

6.1.1 Properties of Cumulative Distribution Functions

(i)

$$x \leq F_X(x) \leq 1 \tag{6.2}$$

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

$$\lim_{x\to\infty} F_X(x) = 1 \tag{6.3}$$

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

$$\lim_{x\to-\infty} F_X(x) = 0 \tag{6.4}$$

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

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

(v) The CDF is continuous from the right.

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

(vi)

$$P[a < X \leq b] = F_X(b) - F_X(a) \tag{6.7}$$

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

$$P[X = b] = F_X(b) - F_X\left(b^-\right) \tag{6.8}$$

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

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

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:

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

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.

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

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.

$$f_X(x) \geq 0 \tag{6.12}$$

(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.

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

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

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

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

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

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:

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

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:

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

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.

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

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

$$\mathbb{E}[c] = c \tag{6.19}$$

(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.

$$\mathbb{E}[X + c] = \mathbb{E}[X] + \mathbb{E}[c] \tag{6.20}$$

(iv) Linearity of Expectations for random variables

$$\mathbb{E}\left[a_0 + a_1 X + a_2 X^2 + \ldots + a_n X^n\right] = a_0 + a_1\,\mathbb{E}[X] + a_2\,\mathbb{E}\left[X^2\right] + \ldots + a_n\,\mathbb{E}[X^n] \tag{6.21}$$

6.4 Variance of Single Continuous Random Variable

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

$$\sigma^2 = \mathrm{VAR}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \tag{6.22}$$

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:

$$\sigma = \mathrm{STD}[X] = \sqrt{\mathrm{VAR}[X]} \tag{6.23}$$

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 $X \sim N\left(\mu, \sigma^2\right)$. μ is $\mathbb{E}[X]$ and σ^2 is how narrow/sharp the bell is. A Gaussian Random Variable has a PDF of:

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

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 Φ.

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

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

6.5.1 Q-Function

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

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

Remark 41.1.

$$Q(Z) = 1 - f_Z(z) \tag{6.27}$$

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:

$$P[X \geq a] \leq \frac{\mathbb{E}[X]}{a} \tag{6.28}$$

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.

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

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

■

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:

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

Proving the Chebychev Inequality.

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

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

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

■

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:

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

- 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:

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

and is calculated as:

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

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 $\{X \le x\} \cap \{Y \le y\}$

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

7.2.1 Properties of Joint Cumulative Distribution Functions

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

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

(ii)

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

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

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

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

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

(v) The probability of the "rectangle" $\{x_1 \le X \le x_2, y_1 \le Y \le y_2\}$

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

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.

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

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 soem nonnegative function $f_{X,Y}(x,y)$, which we call the joint PDF, that is defined on the real plane such that tfor every event B which is a subset of the xy plane

$$P\left[(X,Y)\, inB\right] = \iint_B f_{X,Y}(x,y)\, dxdy \tag{7.11}$$

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

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

$$x \geq 0, y \geq 0 \forall x \forall y \tag{7.13}$$

$$\tag{7.14}$$

7.3.2 Facts about Joint PDFs

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

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

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

$$\tag{7.18}$$

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.

$$\begin{aligned}
f_X(x) &= \frac{d}{dx} F_X(x) \\
&= \frac{d}{dx}\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} f_{X,Y}(s,t)\, dtds\right] \\
&= \frac{d}{dx}\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(s,t)\, dtds \\
&\text{Simplified with Second Fundamental Theorem of Calculus} \\
&= \int_{-\infty}^{\infty} f_{X,Y}(x,t)\, dt \\
f_X &= \int_{-\infty}^{\infty} f_{X,Y}(x,t)\, dt
\end{aligned} \tag{7.19}$$

7.4 Independence of Multiple Random Variables

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

$$P\left[X \in A_1, Y \in A_2\right] = P\left[X \in A_1\right] * P\left[Y \in A_2\right] \tag{7.20}$$

There are 3 ways to phrase this:

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

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

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

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

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

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

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

Independence of Discrete Random Variables with PMF. ■

Theorem 7.1 (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 $Z = g(X,Y)$.

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

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

$$\mathbb{E}[X_1 + X_2 + \ldots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n] \tag{7.25}$$

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

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

7.6 Joint Moments, Correlation, and Covariance

7.6.1 Joint Moments

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

$$\mathbb{E}\left[X^jY^k\right] = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^jy^k\cdot f_{X,Y}(x,y)\,dxdy & \text{if } X,\ Y \text{ are jointly continuous} \\ \sum\limits_{i\in S_X}\sum\limits_{\ell\in S_Y} x_i^jy_l^k\cdot p_{X,Y}(x_i,y_\ell) & \text{if } X,\ Y \text{ are discrete} \end{cases} \tag{7.27}$$

7.6.2 Covariance

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

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

$$\operatorname{Cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \tag{7.29}$$

7.6.3 Correlation

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

Remark 54.1. If X, Y are such that $\mathbb{E}\left[X^1Y^1\right] = 0$, 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

$$\rho_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sigma_X\sigma_Y} \tag{7.30}$$

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 $\operatorname{Cov}[X,Y] = 0$, 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:

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

Remark 56.1. This also implies that

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

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

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

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:

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

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

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

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

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

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:

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

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:

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

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

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

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:

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

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

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

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

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

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

$$\mathbb{E}\left[g\left(x\right)\right] = \mathbb{E}\left[\mathbb{E}\left[Y \,|\, X\right]\right] = \mathbb{E}\left[Y\right] \tag{7.43}$$

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

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

Prove Expectation of Conditional Expected Value. ■

8 Random Vectors

Random Vectors are usually denoted:

$$\vec{X} = \langle X_1, X_2 X_3, \ldots, X_n \rangle \tag{8.1}$$

Defn 62 (Random Vector)**.** A *random vector* is a list of Random Variables.

Remark 62.1. Almost all of the material for Multiple Random Variables is applicable here. However, the 2 random variable equations and definitions must be generalized to n random variables.

8.1 Joint CDF of a Random Vector

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

8.2 Joint PDF of a Random Vector

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

8.2.1 Marginal PDF of a Random Vector

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

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

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

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

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

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

Remark 63.1. This also implies that

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

Remark 63.2. If all elements of $\vec{Y}$ are *independent* (Remember that you need to check each subgroup too, like shown in Section 3.4), then:

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

Remark 63.3. 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

8.4 Mean Vector

Defn 64 (Mean Vector)**.** For $\vec{X} = \langle X_1, X_2, \ldots, X_n \rangle$, the *mean vector* is defined as the column vector of expected values of the components of X_k:

$$\mathbf{m_X} = \mathbb{E}\left[\vec{X}\right] = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix} \tag{8.9}$$

Remark 64.1. Note that we defined the vector of expected values as a column vector. Other texts will use row vectors for other things, but the use of column vectors here is intentional.

8.5 Correlation and Covariance Matrix

Defn 65 (Correlation Matrix)**.** The *correlation matrix* has the second moments of $\bar{X}$ as its entries:

$$\bar{\mathbf{R}}_{\mathbf{X}} = \begin{bmatrix} \mathbb{E}\left[X_1^2\right] & \mathbb{E}[X_1X_2] & \cdots & \mathbb{E}[X_1X_n] \\ \mathbb{E}[X_2X_1] & \mathbb{E}\left[X_2^2\right] & \cdots & \mathbb{E}[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[X_nX_1] & \mathbb{E}[X_nX_2] & \cdots & \mathbb{E}\left[X_n^2\right] \end{bmatrix} \tag{8.10}$$

Remark 65.1. $\bar{R}_X$ is a $n \times n$ symmetric matrix.

Defn 66 (Covariance Matrix)**.** The *covariance matrix* has the second-order central moments as its entries:

$$\begin{aligned} \bar{\mathbf{K}}_{\mathbf{X}} &= \begin{bmatrix} \mathbb{E}\left[(X_1 - m_1)^2\right] & \mathbb{E}[(X_1 - m_1)(X_2 - m_2)] & \cdots & \mathbb{E}[(X_1 - m_1)(X_n - m_n)] \\ \mathbb{E}[(X_2 - m_2)(X_1 - m_1)] & \mathbb{E}\left[(X_2 - m_2)^2\right] & \cdots & \mathbb{E}[(X_2 - m_2)(X_n - m_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[(X_n - m_n)(X_1 - m_1)] & \mathbb{E}[(X_n - m_n)(X_2 - m_2)] & \cdots & \mathbb{E}\left[(X_n - m_n)^2\right] \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{VAR}[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_2, X_1] & \mathrm{VAR}[X_2] & \cdots & \mathrm{Cov}[X_2, X_n] \\ \cdots & \cdots & \ddots & \cdots \\ \mathrm{Cov}[X_n, X_1] & \mathrm{Cov}[X_2, X_n] & \cdots & \mathrm{VAR}[X_n] \end{bmatrix} \end{aligned} \tag{8.11}$$

Remark 66.1. $\bar{K}_X$ is a $n \times n$ symmetric matrix.

Remark 66.2. The diagonal elements of $\bar{K}_X$ are given by the variances $\mathrm{VAR}[X_k] = \mathbb{E}\left[(X_k - m_k)^2\right]$ of the elements of $\vec{X}$.

Remark 66.3. If the diagonal elements of $\bar{K}_X$ are , then $\mathrm{Cov}[X_j, X_k] = 0$ for $j \neq k$, and $\bar{K}_X$, the Covariance Matrix is a diagonal matrix.

Remark 66.4. If the random variables $X_1, X_2, \cdots, X_n$ are independent, then they are uncorrelated and $\bar{K}_X$ is diagonal.

Remark 66.5. If the Mean Vector is $\bar{0}$, that is, $m_k = \mathbb{E}[X_k] = 0$ for all k, then $\bar{R}_X = \bar{K}_X$.

9 Statistics

In applying probability models to real situations, we perform experiments and collect data to answer questions such as:

1. What are the values of the parameters of the distribution of a random variable of interest?
   - Mean or Expected value
   - Variance
2. Is the data set consistent with some model?
   - Some assumed distribution, which must be true, otherwise the model is wrong.
3. Is the data set consistent with some parameter value of the assumed value?

Defn 67 (Random Sample)**.** A *random sample* is a set of n Random Variable or Statistic that are drawn with Independent and Identically Distributed.

$$\mathbf{X}_n = (X_1, X_2, \ldots, X_n) \tag{9.1}$$

Remark 67.1. This is *similar* to the definition of a Random Vector. The difference here is that the values in a Random Sample must be Independent and Identically Distributed and *may* be related to each other somehow.

Remark 67.2 (Random Sample Parameters). These are an additional variable that is added onto the Probability Density Function or Probability Mass Function. When we were using these functions in the previous sections, these parameters were either constant or assumed to be constant. When considering these samples in Statistics, you must also account for the Random Sample Parameters.

Defn 68 (Statistic)**.** A *statistic* $W(\mathbf{X}_n)$ is a function of the random sample $X_1, X_2, \ldots, X_n$.

$$W(\mathbf{X}_n) = g(X_1, X_2, \ldots, X_n) \tag{9.2}$$

Defn 69 (Unit Variance)**.** The *unit variance* means that the standard deviation, σ of a sample, as well as the variance, σ^2 will tend towards 1 as the sample size increases to infinity.

9.1 Sums of Random Variables

Defn 70 (Sum of Random Variables)**.** The definition of a *sum of random variables* is given in Equation (9.3) below. Where X_i is a random variable,

$$S_n = \sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n \tag{9.3}$$

9.1.1 Means and Variances of Sums of Random Variables

Defn 71 (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.*

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] \tag{9.4}$$

Remark 71.1. All the properties of Properties of Expected Values and/or Properties of Expected Value hold true here as well..

Defn 72 (Variance of Sums of Random Variables)**.** The defintion 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.

$$\mathrm{VAR}[S_n] = \mathrm{VAR}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{VAR}[X_i] + \sum_{j=1}^{n} \sum_{\substack{k=1 \\ j \neq k}}^{n} \mathrm{Cov}[X_j, X_k] \tag{9.5}$$

Remark 72.1. If $X_1, X_2, \ldots, X_n$ are independent, then:

$$\mathrm{VAR}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{VAR}[X_i] \tag{9.6}$$

Defn 73 (Independent and Identically Distributed)**.** We say that $X_1, X_2, \ldots, X_n$ are *Independent and Identically Distributed (iid)* random variables if X_i are drawn independently from the same population/probability distribution.

$$\sum_{i=1}^{n} \mathbb{E}[X_i] = n\mu \tag{9.7a}$$

$$\mathrm{VAR}[S_n] = n\sigma^2 \tag{9.7b}$$

- μ is the mean of a random variable X_i
- σ^2 is the variance of a random variable X_i.

9.2 Sample Mean

Defn 74 (Sample Mean)**.** The *sample mean* of a sequence is denoted as,

$$\bar{X} = M_n = \frac{\sum_{i=1}^{n} X_i}{n} \tag{9.8}$$

Defn 75 (Expected Value of Sample Mean)**.** The *expected value of the sample mean* is defined as:

$$\mathbb{E}\left[\bar{X}\right] = \mathbb{E}[M_n] = \frac{\mathbb{E}[S_n]}{n} = \frac{n\mu}{n} = \mu \tag{9.9}$$

Remark 75.1. The sample mean M_n is an *Unbiased Estimator* of population mean μ.

Defn 76 (Variance of Sample Mean)**.** The *variance of the sample mean* is denoted as:

$$\mathrm{VAR}\left[\bar{X}\right] = \mathrm{VAR}\left[M_n\right] = \mathrm{VAR}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\,\mathrm{VAR}\left[S_n\right] = \frac{\sigma^2}{n} \tag{9.10}$$

Remark 76.1. The larger n gets, the smaller $\mathrm{VAR}\left[M_n\right]$ gets, and the closer M_n gets to μ.

Also, we can use the Chebychev Inequality to approximate many values. In this case, we change the Chebychev Inequality from Equation (6.29) to Equation (9.11) like so:

$$P\left[\left|M_n - \mathbb{E}\left[M_n\right]\right| \geq \varepsilon\right] \leq \frac{\mathrm{VAR}\left[M_n\right]}{\varepsilon^2} \tag{9.11}$$

9.3 Important Probability and Statistics Theorems

There are 3 very import theorems that are used quite frequently in both Probability Theory and Statistics.

1. Weak Law of Large Numbers
2. Strong Law of Large Numbers
3. Central Limit Theorem

Theorem 9.1 (Weak Law of Large Numbers)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of Independent and Identically Distributed random variables form a population with mean $\mathbb{E}\left[X\right] = \mu$, then for $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\left[\left|M_n - \mu\right| < \varepsilon\right] = 1 \tag{9.12}$$

Remark. In words this means, for large enough fixed values of n, M_n is close to μ with high probability.

Theorem 9.2 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of Independent and Identically Distributed random variables form a population with mean $\mathbb{E}\left[X\right] = \mu$ and finite variance, then*

$$P\left[\lim_{n\to\infty} M_n = \mu\right] = 1 \tag{9.13}$$

Remark. With probability 1, every sequence of sample mean calculations will eventually approach and stay close to the population mean.

Theorem 9.3 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of Independent and Identically Distributed random variables form a population with mean $\mathbb{E}\left[X\right] = \mu < \infty$ and finite variance σ^2 and let*

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

then,

$$\lim_{n\to\infty} P\left[Z_n \leq z\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\,dx \tag{9.14}$$

Remark. This means that over time, as you gain more and more sample means, they will start to resemble the Gaussian/Normal Random Variable, or the Normal Random Variable.

9.4 Estimators

- A Statistic is a function of the data $X_1, X_2, \ldots, X_n$
- An *estimator* for a parameter, θ, usually denoted $\hat{\theta}$, is also a statistic

Defn 77 (Unbiased Estimator)**.** In general we say that a Statistic $\Theta(X)$ (a function of data $X_1, X_2, \ldots, X_n$) is an *unbiased estimator* of a parameter θ if $\mathbb{E}\left[W\left(X\right)\right] = \theta$.

Remark 77.1 (What makes a good estimator of any parameter, θ?)**.** A *good estimator* of any parameter, θ, should:

- Give the correct value of θ
- Not vary too much around θ

Remark 77.2. This is the definition of *unbiased*, drawn from the definition of Bias

9.4.1 Goodness of an Estimator

There are 4 measures we use to determine how good our estimator is.

1. Bias
2. Variance of Sample Mean
3. Mean Squared Error
4. Consistency

If our estimator is an Unbiased Estimator, then:

- Accuracy is defined as $\mathrm{B}[\hat{\theta}] = \mathbb{E}[\hat{\theta}] - \theta$
- Precision is defined as $\mathrm{VAR}[\hat{\theta}]$

Defn 78 (Bias)**.** *Bias* is defined as:

$$\mathrm{B}[\hat{\Theta}] = \mathbb{E}[\hat{\Theta}] - \theta \tag{9.15}$$

Remark 78.1. The estimator $\hat{\Theta}$ is *unbiased* for θ if

$$\mathbb{E}[\hat{\Theta}] = \theta \tag{9.16}$$

Defn 79 (Mean Squared Error)**.** The *Mean Squared Error* of an estimator for parameter $\hat{\theta}$ is:

$$\mathrm{MSE}[\hat{\theta}] = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right] = \mathrm{VAR}\left[\hat{\theta}\right] + \mathrm{B}^2\left[\hat{\theta}\right] \tag{9.17}$$

Remark. When doing statistical analysis, there is something called the *Bias-Variance Tradeoff*. When doing the analysis, if you try to minimize bias, your variance will increase and vice-versa. There is a happy medium, which is not discussed in this class.

Defn 80 (Consistency)**.** $\hat{\theta}$ is a *consistent estimator* for θ if $\hat{\theta}$ converges to θ in probability.

$$\lim_{n\to\infty} \mathrm{P}\left[|\hat{\theta} - \theta| > \varepsilon\right] = 0 \tag{9.18}$$

9.5 How to Find a Good Estimator

There are several methods, two of which are:

1. Method of Moments
   - Sample Moments and Population Moments, $\bar{X}_n = \mu$
   - You needs as many moments as parameters to get enough equations
2. Maximum Likelihood Estimation

9.5.1 Maximum Likelihood Estimation

Defn 81 (Maximum Likelihood Estimation)**.** Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} f(x \mid \theta)$.

$$\hat{\Theta}_{\mathrm{MLE}} = \operatorname*{argmax}_{\theta\in\Theta} \mathcal{L}(\theta \mid x_1, x_2, \ldots, x_n) \tag{9.19}$$

Remark 81.1. Likelihood, denoted $\mathcal{L}$ is defined as the Joint Probability Density Function of the Random Sample and its Random Sample Parameters.

$$\mathcal{L}(\theta \mid x_1, x_2, \ldots, x_n) = f_{X_1}(x_1 \mid \theta) \cdot f_{X_2}(x_2 \mid \theta) \cdot \ldots \cdot f_{X_n}(x_n \mid \theta) \tag{9.20}$$

Remark 81.2. It is often easier to maximize $\hat{\Theta}_{\mathrm{MLE}}$ over the log-likelihood.

$$\hat{\Theta}_{\mathrm{MLE}} = \operatorname*{argmax}_{\theta\in\Theta} \log \mathcal{L}(\theta \mid x_1, x_2, \ldots, x_n)$$

Remark 81.3.

$$\operatorname*{argmax}_{\theta\in\Theta}$$

is the global maxima of the function. This is further described in Definition A.2.3.

9.6 Confidence Interval

A Reference Material

A.1 Trigonometry

A.1.1 Trigonometric Formulas

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

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

A.1.2 Euler Equivalents of Trigonometric Functions

$$\sin(x) = \frac{e^{\imath x} + e^{-\imath x}}{2} \tag{A.3}$$

$$\cos(x) = \frac{e^{\imath x} - e^{-\imath x}}{2\imath} \tag{A.4}$$

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

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

A.2 Calculus

A.2.1 Fundamental Theorems of Calculus

Defn A.2.1 (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

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

Defn A.2.2 (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

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

Defn A.2.3 (argmax)**.** The arguments to the *argmax* function are to be maximized by using their derivatives. You must take the derivative of the function, find critical points, then determine if that critical point is a global maxima. This is denoted as

$$\operatorname*{argmax}_x$$