Lecture note: Probability and Statistics for EE

Sunghee Yun

July 7, 2025

Contents

Chapter 1

Probability Models in Electrical and Computer Engineering

- Electrical and computer engineers have played a central role in the design of modern information and communications systems. These highly successful systems work reliably and predictably in highly variable and chaotic environments.
- Designers today face even greater challenges. The systems they build are unprecedented in scale and the chaotic environments in which they must operate are untrodden terrritory.
- Probability models are one of the tools that enable the designer to make sense out of the chaos and to successfully build systems that are **efficient, reliable, and cost effective**.

1.1 Mathematical Models as Tools in Analysis and Design

- A **model** is an approximate representation of a physical situation. A model attempts to explain observed behavior using a set of **simple and understandable rules**.
- **Mathematical models** are used when the observational phenomenon has **measurable** properties.

1.2 Deterministic Models

- In **deterministic models** the conditions under which an experiment is carried out determine the **exact outcome** of the experiment. **Circuit theory** is an example of a deterministic mathematical model.

1.3 Probability Models

- Many systems of interest involve phenomena that exhibit unpredictable variation and randomness. We define a **random experiment** to be an experiment in which the outcome varies in an **unpredictable fashion** when the experiment is repeated **under the same conditions**.
- In a random experiment of selecting a ball from an urn containing three identical balls, labeled 0, 1, and 2. The **outcome** of this experiment is a number from the set $S = \{0, 1, 2\}$. We call the set S of all possible outcomes the **sample space**.

- Many probability models in engineering are based on the fact that averages obtained in long sequences of repetitions (trials) of random experiments consistently yield approximately the same value. This property is called **statistical regularity**.

- The modern theory of probability begins with a construction of a set of axioms that specify that probability assignments must satisfy these properties. It supposes that:

  1. a random experiment has been defined, and a set S of all possible outcomes has been identified.
  2. a class of subsets of S called events has been specified
  3. each event A has been assigned a number, $P[A]$, in such a way that the following axioms are satisfied:

  - $0 \leq P[A] \leq 1$.
  - $P[S] = 1$.
  - If A and B are events that cannot occur simultaneously,

- Let us consider how we proceed from a real-world problem that involves randomness to a **probability model** for the problem. The theory requires that we identify the elements in the above axioms. This involves (1) defining the random experiment inherent in the application, (2) specifying the set S of all possible outcomes and the events of interest, and (3) specifying a probability assignment from which the probabilities of all events of interest can be computed. The challenge is to develop the simplest model that explains all the relevant aspects of the real-world problem

1.4 A Detailed Example: A Packet Voice Transmission System

1.5 Other Examples

1.6 Overview of Book Summary Problems

Chapter 2

Basic Concepts of Probability Theory

2.1 Specifying Random Experiments

- A **random experiment** is an experiment in which the outcome **varies in an unpredictable fashion** when the experiment is repeated under the same conditions. A random experiment is specified by stating an experimental procedure and a set of one or more measurements or observations.
- LGExample 2.1
  - Experiment E_1: Select a ball from an urn containing balls numbered 1 to 50. Note the number of the ball.
  - Experiment E_2: Select a ball from an urn containing balls numbered 1 to 4. Suppose that balls 1 and 2 are black and that balls 3 and 4 are white. Note the number and color of the ball you select.
  - Experiment E_3: Toss a coin three times and note the sequence of heads and tails.
  - Experiment E_4: Toss a coin three times and note the number of heads.
  - Experiment E_5: Count the number of voice packets containing only silence produced from a group of N speakers in a 10-ms period.
  - Experiment E_6: A block of information is transmitted repeatedly over a noisy channel until an error-free block arrives at the receiver. Count the number of transmissions required.
  - Experiment E_7: Pick a number at random between zero and one.
  - Experiment E_8: Measure the time between page requests in a Web server.
  - Experiment E_9: Measure the lifetime of a given computer memory chip in a specified environment.
  - Experiment E_{10}: Determine the value of an audio signal at time t_1.
  - Experiment E_{11}: Determine the values of an audio signal at times t_1 and t_2.
  - Experiment E_{12}: Pick two numbers at random between zero and one.
  - Experiment E_{13}: Pick a number X at random between zero and one, then pick a number Y at random between zero and X.
  - Experiment E_{14}: A system component is installed at time $t = 0$. For $t \geq 0$ let $X(t) = 1$ as long as the component is functioning, and let $X(t) = 0$ after the component fails.
- The specification of a random experiment must include an **unambiguous** statement of **exactly what is measured or observed**. For example, random experiments may consist of the same procedure but differ in the observations made, as illustrated by E_3 and E_4.

- The **sample space** S of a random experiment is defined as the set of all possible outcomes. We can

  1. list all the elements, separated by commas, inside a pair of braces:

  $$A = \{0, 1, 2, 3\},$$

  2. give a property that specifies the elements of the set:

  $$A = \{x | x \text{ is an integer such that } 0 \leq x \leq 3\}.$$

- The sample spaces corresponding to the experiments in LGExample 2.1 are given below using set notation:

$$
\begin{aligned}
S_1 &= \{1, 2, \ldots, 50\} \\
S_2 &= \{(1, b), (2, b), (3, w), (4, w)\} \\
S_3 &= \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\} \\
S_4 &= \{0, 1, 2, 3\} \\
S_5 &= \{0, 1, 2, \ldots, N\} \\
S_6 &= \{1, 2, 3, \ldots\} \\
S_7 &= \{x | 0 \leq x \leq 1\} = [0, 1] \\
S_8 &= \{t | t \geq 0\} = [0, \infty) \\
S_9 &= \{t | t \geq 0\} = [0, \infty) \\
S_{10} &= \mathbf{R} = \{v | -\infty < v < \infty\} = (-\infty, \infty) \\
S_{11} &= \mathbf{R}^2 = \{(v_1, v_2) | -\infty < v_1 < \infty \text{ and } -\infty < v_2 < \infty\} \\
S_{12} &= \{(x, y) | 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\} \\
S_{13} &= \{(x, y) | 0 \leq y \leq x \leq 1\} \\
S_{14} &= \text{set of functions } X(t) \text{ for which } X(t) = 1 \text{ for } 0 \leq t < t_0 \text{ and } X(t) = 0 \\
&\quad \text{for } t \geq t_0 \text{ where } t_0 > 0 \text{ is the time when the component fails.}
\end{aligned}
$$

- The three types of sample space:

  1. finite sample space, *e.g.*, E_1, E_2, E_3, E_4, and E_5
  2. discrete sample space or countably infinite sample space, *e.g.*, E_6
  3. continous sample space, *e.g.*, E_7 through E_{13}

- **Events** correspond to subsets of S.

  - For example,
    1. $\{HHH, HHT\}$ is an event of E_3
    2. $[2, \infty)$ is an event of E_8
  - Two special events
    1. S: the sample space itself
    2. $\emptyset$: impossible or null event

- Event examples: A_k refers to an event corresponding to Experiment E_k

  - E_1: "An even-numbered ball is selected," $A_1 = \{2, 4, \ldots, 48, 50\}$.
  - E_2: "The ball is white and even-numbered," $A_2 = \{(4, w)\}$.

- E_3: "The three tosses give the same outcome," $A_3 = \{HHH, TTT\}$.
- E_4: "The number of heads equals the number of tails," $A_4 = \emptyset$.
- E_5: "No active packets are produced," $A_5 = \{0\}$.
- E_6: "Fewer than 10 transmissions are required," $A_6 = \{1, \ldots, 9\}$.
- E_7: "The number selected is nonnegative," $A_7 = S_7$.

- Elementary event: an event from a discrete sample space that consists of a single outcome

Review of Set Theory

- A **set** is a collection of objects and will be denoted by capital letters S, A, B, ... We define U as the **universal set** that consists of all possible objects of interest in a given setting or application. In the context of random experiments we refer to the universal set as the **sample space**.

- A **set** A is a collection of objects from U, and these objects are called the **elements** or **points** of the set A and will be denoted by lowercase letters, ζ, a, b, x, y, ... We use the notation:

$$x \in A \text{ and } x \notin A$$

- A **Venn diagram** is an illustration of sets and their interrelationships.

- We say A is a **subset** of B if every element of A also belongs to B, that is, if $x \in A$ implies $x \in B$. We say that "A is contained in B" and we write:

$$A \subset B$$

or

$$A \subseteq B.$$

- The **empty set** $\emptyset$ is defined as the set with no elements.

- We say that A **and** B **are equal** if they contain the same elements.

$$A = B \text{ if and only if } A \subset B \text{ and } B \subset A.$$

- Three basic operations on sets: **union**, **intersection**, and **complement**
  - union
$$A \cup B = \{x | x \in A \text{ or } x \in B\}$$
  - intersection
$$A \cap B = \{x | x \in A \text{ and } x \in B\}$$
  - complement
$$A^c = \{x | x \notin A\}$$

- **Relative complement** or **difference**

$$A - B = A \backslash B = \{x | x \in A \text{ and } x \notin B\}$$

- Properties of set operations
  - commutativity
$$A \cup B = B \cup A \text{ and } A \cap B = B \cap A.$$

- associativity
$$A \cup (B \cup C) = (A \cup B) \cup C \text{ and } A \cap (B \cap C) = (A \cap B) \cap C$$
- distributivity
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \text{ and } A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

- DeMorgan's rules
$$(A \cup B)^c = A^c \cap B^c \text{ and } (A \cap B)^c = A^c \cup B^c$$
- LGExample 2.7
- The union and intersection operations can be repeated for an arbitrary number of sets.
  - intersection
    * for finite number of sets
    $$\bigcap_{k=1}^{n} A_k = A_1 \cap \cdots \cap A_n$$
    * for countably infinite sequence of sets
    $$\bigcap_{k=1}^{\infty} A_k$$
  - union
    * for finite number of sets
    $$\bigcup_{k=1}^{n} A_k = A_1 \cup \cdots \cup A_n$$
    * for countably infinite sequence of sets
    $$\bigcup_{k=1}^{\infty} A_k$$
- countable unions and intersections of sets are essential in dealing with sample spaces that are not finite.
- **Event class** $\mathcal{F}$ is the set of events of interest.
  - $\mathcal{F}$ is a **set of subsets** of the sample space S
  - The three set operations, *i.e.*, complements, countable unions, and countable intersections, result in events in $\mathcal{F}$
- **Borel field** is the set of subsets in $\mathbf{R}$ which can be generated by the three set operations from intervals of real line. (later...)
- **power set of** S: when $S = \{1, 2, \ldots, k\}$ is a finite sample space, the of all the subsets of S is called the power set of S and it is denoted by 2^S.
- LGExample 2.8

2.2 The Axioms of Probability

- A **probability law** for a random experiment is a rule that assigns probabilities to the events of the experiment that belong to the event class $\mathcal{F}$.

$$P : \mathcal{F} \to \mathbf{R}_+.$$

  For an event $A \in \mathcal{F}$, a number $P[A]$ is called the **probability of** A.

- The probability law must satisfy the following axioms:

  – Axiom I:
$$P[A] \geq 0. \tag{2.1}$$

  – Axiom II:
$$P[S] = 1. \tag{2.2}$$

  – Axiom III: If $A \cap B = \emptyset$, then
$$P[A \cup B] = P[A] + P[B]. \tag{2.3}$$

  – Axiom III': If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, *i.e.*, $A_i \cap A_j = \emptyset$ for all $i \neq j$, then
$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]. \tag{2.4}$$

Corollary 2.1
$$P[A^c] = 1 - P[A]$$

Corollary 2.2
$$P[A] \leq 1$$

Corollary 2.3
$$P[\emptyset] = 0$$

Corollary 2.4 *If $A_1, A_2, \ldots, A_n$ are mutually exclusive, then*
$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P[A_k] \textit{ for } n \geq 2.$$

Corollary 2.5
$$P[A \cup B] = P[A] + P[B] - P[A \cap B].$$

Corollary 2.6
$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P[A_k] - \sum_{j<k}^{n} P[A_j \cap A_k] + \sum_{i<j<k}^{n} P[A_i \cap A_j \cap A_k] \cdots + (-1)^{n+1} P[A_1 \cap \cdots \cap A_n],$$

from which it is follows that
$$P[A \cup B] \leq P[A] + P[B].$$

This results is frequently used to obtain upper bounds for probabilities of interest.

Corollary 2.7 *If $A \subset B$, then*
$$P[A] \leq P[B].$$

However, we still **need an initial probability assignment for some basic set of events** from which the probability of all other events can be computed.

Discrete Sample Spaces

- The **probability law** for an experiment with a countable sample space can be specified by giving the **probabilities of the elementary events**.

- **Finite case:** Suppose that $S = \{a_1, a_2, \ldots, a_n\}$ and let $\mathcal{F}$ consist of all subsets of S. Then for any event $B = \{a'_1, a'_2, \ldots, a'_m\} \in \mathcal{F}$,

$$P[B] = P[\{a'_1, a'_2, \ldots, a'_m\}] = P[\{a'_1\}] + P[\{a'_1\}] + \cdots + P[\{a'_m\}]$$

  - Of particular interest is the case of **equally likely outcomes**.

$$P[\{a_1\}] = P[\{a_2\}] = \cdots = P[\{a_n\}] = \frac{1}{n}.$$

    If $B = \{a'_1, a'_2, \ldots, a'_k\} \in \mathcal{F}$,

$$P[B] = P[\{a'_1, a'_2, \ldots, a'_k\}] = P[\{a'_1\}] + P[\{a'_1\}] + \cdots + P[\{a'_k\}] = \frac{k}{n},$$

    i.e., if outcomes are equally likely, then the probability of an event is equal to the number of outcomes in the event divided by the total number of outcomes in the sample space.

- **Countably infinite case:** Suppose $S = \{a_1, a_2, \ldots\}$. Axiom III' implies that if $D = \{a'_1, a'_2, \ldots\}$,

$$P[D] = P[\{a'_1, a'_2, a'_3, \ldots\}] = P[\{a'_1\}] + P[\{a'_2\}] + P[\{a'_3\}] + \cdots,$$

  i.e., the probability of an event with a countably infinite sample space is determined from the infinite (or finite) sum of the probabilities of the elementary events.

- LGExample 2.10 a coin is tossed three times.

- LGExample 2.11 a coin is tossed repeatedly until the first heads shows up.

Continuous Sample Spaces

- All the subset of $\mathbf{R}$ is too large!

- The **Borel field**, $\mathcal{B}$, contains all open and closed intervals of the real line as well as all events that can be obtained as countable unions, intersections, and complements. **Axiom III' is once again the key to calculating probabilities of events**, *i.e.*,

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$$

  for a sequence of mutually exclusive events that are represented by intervals of the real line. For this reason, probability laws in experiments with continuous sample spaces specify a rule for assigning numbers to **intervals of the real line**.

- **The probability that the outcome takes on a specific value is zero.** Why?

- LGExample 2.13 The proportion of chips whose lifetime exceeds t decreases exponentially at a rate α. Then

$$P[(t, \infty)] = e^{-\alpha t}$$

  for $t > 0$.

How do we proceed from a problem statement to its probability model?

1. The problem statement implicitly or explicitly **defines a random experiment**, which specifies an experimental procedure and a set of measurements and observations. These **measurements and observations determine the set of all possible outcomes** and hence the sample space S.
2. An **initial probability** assignment that specifies the probability of certain events must be determined next. This probability assignment must satisfy the axioms of probability.
   - If S is **discrete**, then it suffices to specify the **probabilities of elementary events**.
   - If S is **continuous**, it suffices to specify the **probabilities of intervals of the real line or regions of the plane**.
3. The probability of other events of interest can then be determined from the **initial probability assignment** and **the axioms of probability** and their corollaries.

2.3 Computing Probabilities Using Counting Methods †

This section is optional. However, reading it carefully will make it easier for you to understand subsequent sections; I encourage you to read it if you have time.

This section deals with

- permutations
- n factorial: $n! = 1 \cdot 2 \cdots (n-1)n$
- Stirling's formula
$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n} \tag{2.5}$$
- binomial coefficient:
$$\binom{n}{k} = C_k^n = \frac{n!}{k!(n-k)!} \tag{2.6}$$
  for $0 \le k \le n$ with the convention $0! = 1$.
- multinomial coefficient:
$$\frac{n!}{k_1! \cdots k_m!} \tag{2.7}$$
  where $k1 + \cdots + k_m = n$.

2.4 Conditional Probability

- **Conditional probability**, $P[A|B]$, is the probability of event A given that even B has occurred.
$$P[A|B] = \frac{P[A \cap B]}{P[B]} \text{ for } P[B] > 0. \tag{2.8}$$
  This implies
$$P[A \cap B] = P[A|B]P[B] \tag{2.9}$$
  and by symmetry
$$P[A \cap B] = P[B|A]P[A]. \tag{2.10}$$
- Please read and understand (thoroughly) LGExample 2.24 and LGExample 2.25.
- LGExample 2.26 **Binary Communication System**

- Let $B_1, \ldots, B_n$ be mutually exclusive events whose union equals the sample space S as shown in LGFig 2.12, *i.e.*,

$$\bigcup_{k=1}^{n} B_k = S \text{ and } B_i \cap B_j = \emptyset \text{ for all } i \neq j. \tag{2.11}$$

  We refer to these sets as a **partition** of S.

- Any event A can be represented as the union of mutually exclusive events in the following way:

$$A = A \cap S = A \cap \bigcup_{k=1}^{n} B_k = \bigcup_{k=1}^{n} (A \cap B_k)$$

  by the distributivity. Corollary 2.4 implies that

$$P[A] = \sum_{k=1}^{n} P[A \cap B_k] = P[A \cap B_1] + \cdots + P[A \cap B_n]. \tag{2.12}$$

- **Theorem on total probability**: (2.12) and (2.9) imply

$$P[A] = \sum_{k=1}^{n} P[A|B_k]P[B_k] = P[A|B_1]P[B_1] + \cdots + P[A|B_n]P[B_n]. \tag{2.13}$$

- LGExample 2.28

Bayes' Rule

- Let $B_1, \ldots, B_n$ be a partition of a sample space S. Suppose that event A occurs. Then the probability of event B_j is

$$P[B_j|A] = \frac{P[B_j \cap A]}{P[A]} = \frac{P[B_j|A]P[A]}{\sum_{k=1}^{n} P[A|B_k]P[B_k]}. \tag{2.14}$$

  where the theorem on total probability is used and (2.14) is called **Bayes' rule**.

- Bayes' rule is often applied in the following situation. We have some random experiment in which the events of interest form a partition. The **"a priori probabilities"** of these events, $P[B_j]$, are the probabilities of the events before the experiment is performed. Now suppose that the experiment is performed, and we are informed that event A occurred; the **"a posteriori probabilities"** are the probabilities of the events in the partition, $P[B_j|A]$, given this additional information.

- LGExample 2.30 **Quality Control**

2.5 Independence of Events

- If knowledge of the occurrence of an event B does not alter the probability of some other event A, then it would be natural to say that event A is independent of B. In terms of probabilities this situation occurs when

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

  We will define two events A and B to be **independent** if

$$P[A \cap B] = P[A]P[B]. \tag{2.15}$$

Note that if $P[B] \neq 0$, (2.15) with (2.9) implies

$$P[A|B] = P[A] \tag{2.16}$$

and if $P[A] \neq 0$, (2.15) with (2.10) implies

$$P[B|A] = P[B]. \tag{2.17}$$

- In general if two events have nonzero probability and are mutually exclusive, then they cannot be independent. For suppose they were independent and mutually exclusive; then

  $$0 = P[A \cap B] = P[A]P[B],$$

  which implies that at least one of the events must have zero probability.

- LGExample 2.32

- **Independence of three events** A, B, and C:

  – pairwise independence:

  $$P[A \cap B] = P[A]P[B], P[B \cap C] = P[B]P[C], \text{ and } P[C \cap A] = P[C]P[A].$$

  – The knowledge of the joint occurrence of any two, say A and B, should not affect the probability of the third, *i.e.*,

  $$P[C] = P[C|A \cap B] \iff P[C] = \frac{P[A \cap B \cap C]}{P[A \cap B]} = \frac{P[A \cap B \cap C]}{P[A]P[B]},$$

  hence

  $$P[A \cap B \cap C] = P[A]P[B]P[C].$$

  – **Independence of** n **events**: The events $A_1, \ldots, A_n$ are said to be independent if for any $k = 2, \ldots, n$ and any $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

  $$P[A_{i_1} \cap \cdots \cap A_{i_k}] = P[A_{i_1}] \cdots P[A_{i_k}]. \tag{2.18}$$

  We must check $2^n - n - 1$ combinations!

  – Most common application of the independence concept is in making the assumption that the events of separate experiments are independent. We refer to such experiments as **independent experiments**.

2.6 Sequential Experiments

- Many random experiments can be viewed as sequential experiments that consist of a sequence of simpler subexperiments.

Sequences of Independent Experiments

- Suppose that a **random experiment consists of performing experiments** $E_1, \ldots, E_n$. The outcome of this experiment will then be an n**-tuple**

  $$s = (s_1, ..., s_n),$$

  where s_k is the outcome of the kth subexperiment. The sample space of the sequential experiment is defined as the set that contains the above n-tuples and is denoted by the **Cartesian product** of the individual sample spaces

  $$S_1 \times \cdots \times S_n. \tag{2.19}$$

- We can usually determine, **because of physical considerations,** when the subexperiments are independent, in the sense that the outcome of any given subexperiment cannot affect the outcomes of the other subexperiments. Let A_1, ..., A_n be events such that A_k concerns only the outcome of the kth subexperiment. **If the subexperiments are independent, then it is reasonable to assume that the above events A_1, ..., A_n are independent.** Thus

$$P[A_1 \cap \cdots \cap A_n] = P[A_1] \cdots P[A_n].$$

This expression allows us to compute all probabilities of events of the sequential experiment.

The Binomial Probability Law

- A **Bernoulli** trial involves performing an experiment once and noting whether a particular event A occurs. The outcome of the Bernoulli trial is said to be a "success" if A occurs and a "failure" otherwise.
- Let k be the number of successes in n independent Bernoulli trials, then the probabilities of k are given by the **binomial probability law:**

$$p_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 0, \ldots, n, \tag{2.20}$$

where $p_n(k)$ is the probability of k successes in n trials, and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{2.21}$$

is the **binomial coefficient**.

- The **binomial theorem** implies

$$(a+b)^n = \sum_{k=1}^{n} \binom{n}{k} a^k b^{n-k}, \tag{2.22}$$

hence

$$\sum_{k=1}^{n} p_n(k) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$$

confirming that **the binomial probabilities sum to** 1.

- recursive formula for the binomial probability:

$$p_n(k+1) = \frac{(n-k)p}{(k+1)(1-p)} p_n(k). \tag{2.23}$$

- LGExample 2.40 **Error Correction Coding**

The Multinomial Probability Law

- Let B_1, B_2, ..., B_M be a partition of the sample space S of some random experiment and let $P[B_j] = p_j$. The events are **mutually exclusive**, so

$$p_1 + p_2 + ... + p_M = 1.$$

Suppose that n independent repetitions of the experiment are performed. Let k_j be the number of times event B_j occurs, then the vector $(k_1, k_2, \ldots, k_M)$ specifies the number of times each of the events B_j occurs. The probability of the vector $(k_1, k_2, \ldots, k_M)$ satisfies the **multinomial probability law:**

$$P[(k_1, k_2, \ldots, k_M)] = \frac{n!}{k_1! k_2! \cdots k_M!} p_1^{k_1} p_2^{k_2} \cdots p_M^{k_M} \tag{2.24}$$

where $k_1 + k_2 + \cdots + k_M = n$. It reduces to the binomial probability law when $M = 2$.

The Geometric Probability Law

- Consider a sequential experiment in which we **repeat independent Bernoulli trials until the occurrence of the first success**. Let the outcome of this experiment be m, the number of trials carried out until the occurrence of the first success. Then

$$p(m) = P[{A_1}^c {A_2}^c \cdots {A_{m-1}}^c A_m] = (1-p)^{m-1}p \text{ for } m \in \{1, 2, \ldots\}. \tag{2.25}$$

  This is called the **geometric probability law**.

- The formula for the sum of infinite geometric sequence gives

$$\sum_{m=1}^{\infty} p(m) = \sum_{m=1}^{\infty} pq^{m-1} = \frac{p}{1-q} = 1$$

  where $q = 1 - p$. Also

$$P[m > K] = \sum_{m=K+1}^{\infty} p(m) = \sum_{m=K+1}^{\infty} pq^{m-1} = \frac{pq^k}{1-q} = q^k.$$

- LGExample 2.43 **Error Control By Retransmission**

Sequences of Dependent Experiments

- Suppose the sequence of outcomes $s_1, s_2, \ldots$ satisfy

$$P[s_n | s_1 \cap s_2 \cap \cdots \cap s_{n-1}] = P[s_n | s_{n-1}] \tag{2.26}$$

  for any $n \leq 1$, *i.e.*, the outcome s_n only depends on the previous outcome s_{n+1}. The sequential experiments that satisfy (2.26) are called **Markov chains** For these experiments, the probability of a sequence $s_1, s_2, \ldots, s_n$ is given by

$$P[s_1, s_2, \ldots, s_n] = P[s_n|s_{n-1}]P[s_{n-1}|s_{n-2}] \cdots P[s_1|s_0]P[s_0]. \tag{2.27}$$

2.7 Synthesizing Randomness: Random Number Generators †

MAYBE LATER

2.8 Fine Points: Event Classes †

NO CONTENTS

2.9 Fine Points: Probabilities of Sequences of Events Summary Problems †

(ALMOST) NO CONTENTS

2.9.1 The Borel Field of Events

2.9.2 Continuity of Probability

Corollary 2.8 (Continuity of Probability Function) *Let A_1, A_2, $\ldots$ be an increasing or decreasing sequences of events in $\mathcal{F}$, then:*

$$\lim_{n\to\infty} P[A_n] = P[\lim_{n\to\infty} A_n]. \tag{2.28}$$

Chapter 3

Discrete Random Variables

3.1 The Notion of a Random Variable

- The outcome of a random experiment **need not be a number**. However, we are usually interested not in the outcome itself, but rather in **some measurement or numerical attribute** of the outcome.
- A **random variable** X is a function that assigns a real number, $X(\zeta)$, to each outcome ζ in the sample space of a random experiment. (LGFig 3.1)
- The sample space S is the **domain** of the random variable, and the set S_X of all values taken on by X is the **range** of the random variable.
- Capital letters denote random variables, *e.g.*, X or Y, and lower case letters denote possible values of the random variables, *e.g.*, x or y
- LGExample 3.1 **Coin Tosses**
- For random variables, the function or rule that assigns values to each outcome is **fixed and deterministic**. Therefore the distribution of the values of a random variable X is determined by the probabilities of the outcomes ζ in the random experiment. In other words, the randomness in the observed values of X is induced by the **underlying random experiment**.

$$P[X \in B] = P[\{\zeta | X(\zeta) \in B\}]$$

3.2 Discrete Random Variables and Probability Mass Function

- A **discrete random variable** X is defined as a random variable that assumes values from a **countable set**, that is, $S_X = \{x_1, x_2, x_3, \ldots\}$. A discrete random variable is said to be **finite** if its range is finite, that is, $S_X = x_1, x_2, \ldots, x_n$.
- We **only need to obtain the probabilities for the events** $A_k = \{\zeta | X(\zeta) = x_k\}$.
- The probability mass function (PMF) of a discrete random variable X is defined as:

$$p_X(x_k) = P[X = x_k] = P[A_k] \text{ for } x \in \mathbf{R}. \tag{3.1}$$

- The events $A_1, A_2, \ldots$ form a **partition** of S as illustrated in LGFig 3.3.

  Proof:

– For $j \neq k$,

$$A_j \cap A_k = \{\zeta | X(\zeta) = x_j \text{ and } X(\zeta) = x_k\} = \emptyset$$

since $X(\zeta)$ cannot have two different values; it's a function.

– For each $\zeta \in S$, $X(\zeta) = x_k$ for some k since $x_1, x_2, \ldots$ is a range of the function $X(\zeta)$. Hence

$$S = \bigcup_{k=1}^{\infty} A_k.$$

- The PMF $p_X(x)$ satisfies three properties that provide all the information required to calculate probabilities for events involving the discrete random variable X:

  1. Axiom I implies
  $$p_X(x) = P[\{\zeta | X(\zeta) = x\}] \geq 0 \; \forall x. \tag{3.2}$$
  2. Axiom II implies
  $$\sum_{x \in S_X} p_X(x) = \sum_{x \in S_X} P[\{\zeta | X(\zeta) = x\}] = \sum_{k=1}^{\infty} P[A_k] = P[S] = 1. \tag{3.3}$$
  3. Axiom III' implies
  $$P[X \in B] = P\left[\bigcup_{x \in B} \{\zeta | X(\zeta) = x\}\right] = P\left[\bigcup_{x_k \in B} A_k\right] = \sum_{x_k \in B} P[A_k] = \sum_{x_k \in B} p_X(x_k). \tag{3.4}$$

- The PMF of X gives us the probabilities for **all the elementary events** from S_X.

- If we are only interested in events concerning X, then we can forget about the underlying random experiment and its associated probability law and **just work with S_X and the PMF of X**.

- LGExample 3.5 **Coin Tosses and Binomial Random Variable**

- LGFig 3.4(a) and LGFig 3.4(b) show the graph of $p_X(x)$ versus x for the random variables. In general, the graph of the PMF of a discrete random variable has **vertical arrows** of height $p_X(x_k)$ at the values x_k in S_X.

- The **uniform random variable**: $p_X(k) = 1/M$ for $k \in \{1, 2, \ldots, M\}$.

- The **Bernoulli random variable** I_A is equal to 1 if A occurs and zero otherwise, and is given by the indicator function for A:
$$I_A(\zeta) = \begin{cases} 1 & \text{if } \zeta \in A, \\ 0 & \text{if } \zeta \notin A. \end{cases}$$
If $P[A] = p$,
$$p_I(x) = \begin{cases} P[\{\zeta | \zeta \in A^c\}] = p & \text{if } x = 1 \\ P[\{\zeta | \zeta \in A\}] = 1 - p & \text{if } x = 0. \end{cases} \tag{3.5}$$

- LGExample 3.9 **Message Transmission** X is a discrete random variable taking on values from $S_X = \{1, 2, 3, \ldots\}$. The event $\{X = k\}$ occurs if the underlying experiment finds $k-1$ consecutive erroneous transmissions ("failures") followed by an error-free one ("success"):
$$p_X(k) = (1-p)^{k-1} p \text{ for } k = 1, 2, \ldots \tag{3.6}$$
We call X the **geometric random variable**, and we say that X is **geometrically distributed**.

- LGExample 3.10 **Transmission Errors** A binary communications channel introduces a bit error in a transmission with probability p. Let X be the number of errors in n independent transmissions. Find the PMF of X. Find the probability of one or fewer errors.

$$p_X(k) = P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 1, 2, \ldots \tag{3.7}$$

  We call X the **binomial random variable**.

- The graph of relative frequencies should approach the graph of the PMF. (LGFig 3.5)

- **the relationship between relative frequencies and the PMF** $p_X(x_k)$. Suppose we perform n **independent repetitions** to obtain n observations of the discrete random variable X. Let $N_k(n)$ be the number of times the event $X = x_k$ occurs and let $f_k(n) = N_k(n)/n$ be the corresponding relative frequency. As n becomes large we expect that $f_k(n) \to p_X(x_k)$. Therefore **the graph of relative frequencies should approach the graph of the PMF**. (LGFig 3.5(a))

3.3 Expected Value and Moments of Discrete Random Variable

- In some situations we are interested in a few parameters that summarize the information provided by the PMF. (LGFig 3.6)

- The **expected value** or **mean** of a discrete random variable X is defined by

$$m_X = \mathbf{E}[X] = \sum_{x \in S_X} x p_X(x) = \sum_k x_k p_X(x_k). \tag{3.8}$$

  The expected value $\mathbf{E}[X]$ is defined if the above sum converges absolutely, that is,

$$\mathbf{E}[|X|] = \sum_k |x_k| p_X(x) < \infty. \tag{3.9}$$

- If we view $p_X(x)$ as the distribution of mass on the points $x_1, x_2, \ldots$ in the real line, then $\mathbf{E}[X]$ represents the **center of mass of this distributiona**.

- LGExample 3.11 **Mean of Bernoulli Random Variable**

$$\mathbf{E}[I_A] = 0 p_I(0) + 1 p_I(1) = p.$$

- The use of the term "expected value" **does not mean that we expect to observe** $\mathbf{E}[X]$ when we perform the experiment that generates X.

- The arithmetic average, or **sample mean**, of the observations, is

$$\begin{aligned} \langle X \rangle_n &= \frac{x(1) + x(2) + \cdots + x(n)}{n} = \frac{x_1 N_1(n) + x_2 N_2(n) + \cdots}{n} \\ &= x_1 f_1(x) + x_2 f_2(x) + \cdots = \sum_k x_k f_k(n). \end{aligned} \tag{3.10}$$

  As n becomes large, we expect relative frequencies to approach the probabilities $p_X(x_k)$:

$$\lim_{n \to \infty} f_k(n) \to p_X(x_k) \text{ for all } k,$$

  hence

$$\langle X \rangle_n = \sum_k x_k f_k(n) \to \sum_k x_k p_X(x_k) = \mathbf{E}[X].$$

- LGExample 3.15 **Mean of a Geometric Random Variable** Let X be the number of bytes in a message, and suppose that X has a geometric distribution with parameter p. Find the mean of X.

  By definition,
$$\mathbf{E}[X] = \sum_{k=1}^{\infty} kpq^{k-1},$$
  thus (A.6) implies
$$\mathbf{E}[X] = p\sum_{k=1}^{\infty} kq^{k-1} = p/(1-q)^2 = 1/p.$$

- LGExample 3.16 **St. Petersburg Paradox** A fair coin is tossed repeatedly until a tail comes up. If X tosses are needed, then the casino pays the gambler $Y = 2^X$ dollars. How much should the gambler be willing to pay to play this game?

  – Random variables with **unbounded expected value** are not uncommon and appear in models where outcomes that have extremely large values are not that rare. Examples include the sizes of files in **Web transfers, frequencies of words in large bodies of text, and various financial and economic problems**.

- Expected value is a **linear operator**: Assume two random variables X and Y and two real numbers a and b. Then
$$\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]. \tag{3.11}$$

3.3.1 Expected Value of Functions of a Random Variable

- Let X be a discrete random variable, and let $Z = g(X)$. Since X is discrete, $Z = g(X)$ will assume a **countable set of values** of the form $g(x_k)$ where $x_k \in S_X$. Denote the set of values assumed by $g(X)$ by $\{z_1, z_2, \ldots\}$. One way to find the expected value of Z is to use (3.8), which requires that we first find the PMF of Z.
$$\mathbf{E}[Z] = \sum_j z_j p_Z(z_j). \tag{3.12}$$
  Another way is to use the following result:
$$\mathbf{E}[Z] = \mathbf{E}[g(X)] = \sum_k g(x_k)p_X(x_k). \tag{3.13}$$
  To show the equivalence of (3.12) and (3.13), we group the terms x_k that are mapped to each value z_j:
$$\sum_k g(x_k)p_X(x_k) = \sum_j z_j \left(\sum_{x_k : g(x_k) = z_j} p_X(x_k) \right) = \sum_j z_j p_Z(z_j) = \mathbf{E}[Z].$$

- LGExample 3.17 **Square-Law Device** Let X be a noise voltage that is uniformly distributed in $S_X = \{-3, -1, +1, +3\}$ with $p_X(k) = 1/4$ for k in S_X. Find $\mathbf{E}[Z]$ where $Z = X^2$.

3.3.2 Variance of a Random Variable

- We are interested not only in the mean of a random variable, but also in the **extent of the random variable's variation** about its mean.

- The **variance of the random variable** X is defined as the expected value of the square of the deviation about the mean, $(X - \mathbf{E}[X])^2$:

$$\sigma_X^2 = \mathbf{VAR}[X] = \mathbf{E}[(X - m_X)^2] = \sum_{x \in S_X} (x - m_X)^2 p_X(x) = \sum_k (x_k - m_X)^2 p_X(x_k). \tag{3.14}$$

  An alternative expression for the variance:

$$\mathbf{E}[(X - m_X)^2] = \mathbf{E}[X^2 - 2m_X X + m_X^2] = \mathbf{E}[X^2] - 2m_X \mathbf{E}[X] + m_X^2 = \mathbf{E}[X^2] - m_X^2 \tag{3.15}$$

- The n**th moment of** X is defined by

$$\mathbf{E}[X^n] \tag{3.16}$$

  - the **second moment of** X is $\mathbf{E}[X^2]$.

- The **standard deviation of the random variable** X is defined by the square root of the variance of X:

$$\sigma_X = \mathbf{STD}[X] = \mathbf{VAR}[X]^{1/2} \tag{3.17}$$

- **Neither** the variance **nor** the standard deviation is a linear operator, but
  - **Adding a constant** to a random variable does not affect the variance.

$$\mathbf{VAR}[X + c] = \mathbf{E}[(X + c - (\mathbf{E}[X] + c))^2] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{VAR}[X].$$

  - **Scaling a random variable** by c scales the variance by c^2 and the standard deviation by $|c|$:

$$\mathbf{VAR}[cX] = \mathbf{E}[(cX - c\mathbf{E}[X])^2] = \mathbf{E}[c^2(X - \mathbf{E}[X])^2] = c^2\mathbf{E}[(X - \mathbf{E}[X])^2] = c^2\mathbf{VAR}[X],$$

    thus

$$\mathbf{STD}[cX] = |c|\ \mathbf{STD}[X]$$

- LGExample 3.22 **Variance of Geometric Random Variable** Find the mean and variance of the geometric random variable.

  The infinite series formula (A.6) implies

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} kq^{k-1}p = p/(1-q)^2 = 1/p. \tag{3.18}$$

  The second moment of X is

$$\mathbf{E}[X^2] = \sum_{k=1}^{\infty} k^2 q^{k-1} p = (1+q)p/(1-q)^3 = (1+q)/p^2 \tag{3.19}$$

  where (A.11) is used, thus the variance is

$$\mathbf{VAR}[X] = \mathbf{E}[X^2] - m_X^2 = (1+q)/p^2 - 1/p^2 = q/p^2 = (1-p)/p^2. \tag{3.20}$$

3.4 Conditional Probability Mass Function

3.4.1 Conditional Probability Mass Function

- The **conditional probability mass function** of X is defined by the conditional probability:

$$p_X(x|C) = P[X = x|C] \tag{3.21}$$

  or equivalently

$$p_X(x|C) = \frac{P[\{X = x\} \cap C]}{P[C]} \tag{3.22}$$

  The above expression has a nice intuitive interpretation: The conditional probability of the event $\{X = x_k\}$ is given by the probabilities of outcomes ζ for which **both** $X(\zeta) = x_k$ **and** ζ **are in** C**, normalized by** $P[C]$.

- The conditional PMF satisfies properties like (3.2), (3.3), and (3.4):

  1. $p_X(x_k|C) \geq 0$ for all $x_k \in S_X$.
  2. $\sum_{x_k \in S_X} p_X(x_k|C) = 1$.
  3. $P[X \in B|C] = \sum_{x_k \in B} p_X(x_k|C)$ where $B \subset S_X$.

  Proof: It is obvious that $p_X(x_k|C) \geq 0$ for all $x_k \in S_X$. Since the set of the events $A_k = \{\zeta | X(\zeta) = x_k\}$ is a partition of S,

$$\sum_{x \in S_X} p_X(x|C) = \sum_{x \in S_X} P[X = x|C] = \sum_{x \in S_X} P[\{X = x\} \cap C]/P[C] = P[C]/P[C] = 1.$$

  Since A_k are mutually exclusive, $A_k \cap C$ are mutually exclusive and

$$\begin{aligned} P[\{X \in B\} \cap C] &= P\left[\left(\bigcup_{x_k \in B} A_k\right) \cap C\right] \\ &= P\left[\bigcup_{x_k \in B} (A_k \cap C)\right] = \sum_{x_k \in B} P[A_k \cap C] = \sum_{x_k \in B} P[\{X = x_k\} \cap C]. \end{aligned}$$

  Thus

$$P[X \in B|C] = \frac{P[\{X \in B\} \cap C]}{P[C]} = \sum_{x_k \in B} \frac{P[\{X = x_k\} \cap C]}{P[C]} = \sum_{x_k \in B} p_X(x_k|C).$$

- LGExample 3.24 **Residual Waiting Times** Let X be the time required to transmit a message, where X is a uniform random variable with $S_X = \{1, 2, \ldots, L\}$. Suppose that a message has already been transmitting for m time units, find the probability that the remaining transmission time is j time units.

  We are given $C = \{X > m\}$, so for $m + 1 \leq m_j \leq L$,

$$p_X(m + j|X > m) = \frac{P[X = m + j]}{P[X > m]} = \frac{1/L}{1 - m/L} = \frac{1}{L - m}$$

  for $1 \leq j \leq L - m$.

- Let $B_1, B_2, \ldots, B_n$ be a partition for the sample space S. The theorem on total probability allows us to find the **PMF of** X **in terms of the conditional PMF's**:

$$p_X(x) = \sum_{i=1}^{n} p_X(x|B_i) P[B_i]. \tag{3.23}$$

Proof: The theorem on total probability implies

$$P[A] = \sum_{k=1}^{n} P[A|B_k]P[B_k]$$

where $A = \{\zeta | X(\zeta) = x\}$.

- LGExample 3.25 **Device Lifetimes** A production line yields two types of devices. Type 1 devices occur with probability α and work for a relatively short time that is geometrically distributed with parameter r. Type 2 devices work much longer, occur with probability $1-\alpha$, and have a lifetime that is geometrically distributed with parameter s. Let X be the lifetime of an arbitrary device. Find the PMF of X.

  The conditional probabilities can be found as

  $$p_X(k|B_1) = (1-r)^{k-1} r \text{ for } k \in \mathbf{N}$$

  and

  $$p_X(k|B_2) = (1-s)^{k-1} s \text{ for } k \in \mathbf{N}.$$

  Thus,

  $$p_X(k) = p_X(k|B_1)P[B_1] + p_X(k|B_2)P[B_2] = (1-r)^{k-1} r\alpha + (1-s)^{k-1} s(1-\alpha).$$

3.4.2 Conditional Expected Value

- The **conditional expected value of** X **given** B is defined as:

  $$m_{X|B} = \mathbf{E}[X|B] = \sum_{x \in S_X} x p_X(x|B) = \sum_k x_k p_X(x_k|B) \tag{3.24}$$

  where we apply the absolute convergence requirement on the summation

- The **conditional variance of** X **given** B is defined as:

  $$\mathbf{VAR}[X|B] = \mathbf{E}[(X - m_{X|B})^2|B] = \sum_k (x_k - m_{X|B})^2 p_X(x_k|B) = \mathbf{E}[X^2|B] - m_{X|B}^2 \tag{3.25}$$

  Note that the variation is **measured with respect to** $m_{X|B}$**, not** m_X.

- Let $B_1, B_2, \ldots, B_n$ be a partition for the sample space S, and let $p_X(x|B_i)$ be the conditional PMF of X given event B_i. $\mathbf{E}[X]$ **can be calculated from the conditional expected values** $\mathbf{E}[X|B]$:

  $$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X|B_i]P[B_i]. \tag{3.26}$$

  Proof: The theorem on total probability implies

  $$\begin{aligned} \mathbf{E}[X] &= \sum_k x_k p_X(x_k) = \sum_k x_k \left(\sum_{i=1}^{n} p_X(x_k|B_i)P[B_i] \right) = \sum_{i=1}^{n} \left(\sum_k x_k p_X(x_k|B_i) \right) P[B_i] \\ &= \sum_{i=1}^{n} \mathbf{E}[X|B_i]P[B_i]. \end{aligned}$$

- Using the same approach we can also show

  $$\mathbf{E}[g(X)] = \sum_{i=1}^{n} \mathbf{E}[g(X)|B_i]P[B_i]. \tag{3.27}$$

Proof: The theorem on total probability implies

$$\begin{aligned}
\mathbf{E}[g(X)] &= \sum_k g(x_k)p_X(x_k) = \sum_k g(x_k)\left(\sum_{i=1}^n p_X(x_k|B_i)P[B_i]\right) \\
&= \sum_{i=1}^n \left(\sum_k g(x_k)p_X(x_k|B_i)\right)P[B_i] = \sum_{i=1}^n \mathbf{E}[g(X)|B_i]P[B_i].
\end{aligned}$$

- LGExample 3.26 **Device Lifetimes** Find the mean and variance for the devices in LGExample 3.25.

  Note that (3.18) and (3.19) imply

  $$\begin{aligned}
  \mathbf{E}[X|B_1] = 1/r, \quad \mathbf{E}[X^2|B_1] = (2-r)/r^2, \\
  \mathbf{E}[X|B_2] = 1/s, \quad \mathbf{E}[X^2|B_2] = (2-s)/s^2.
  \end{aligned}$$

  Now (3.26) and (3.27) imply that

  $$\mathbf{E}[X] = \mathbf{E}[X|B_1]\alpha + \mathbf{E}[X|B_2](1-\alpha) = \alpha/r + (1-\alpha)/s$$

  and

  $$\begin{aligned}
  &\mathbf{VAR}[X] = \mathbf{E}[X^2] - m_X^2 \\
  &= \mathbf{E}[X^2|B_1]\alpha + \mathbf{E}[X^2|B_2](1-\alpha) - m_X^2 \\
  &= \alpha(2-r)/r^2 + (1-\alpha)(2-s)/s^2 - (\alpha/r + (1-\alpha)/s)^2.
  \end{aligned}$$

  (I think the author made a mistake on this.)

  - Note that we do **not** use the conditional variances to find $\mathbf{VAR}[Y]$ because (3.27) does not apply to (conditional) variances. However, the equation **does** apply to the conditional second moments.

3.5 Important Discrete Random Variables

- Certain random variables arise in many diverse, unrelated applications. The pervasiveness of these random variables is due to the fact that **they model fundamental mechanisms** that underlie random behavior.
- **Discrete random variables** arise mostly in applications where **counting is involved**.

3.5.1 The Bernoulli Random Variables

- The Bernoulli random variable I_A equals one if the event A occurs, and zero otherwise.
  - range:

    $$S_X = \{0, 1\}. \tag{3.28}$$

  - PMF:

    $$p_I(0) = 1-p,\ p_I(1) = p. \tag{3.29}$$

    where $P[A] = p$. Figure 3.1 shows the PMF when $p = 0.6$.
  - mean and variance:

    $$\mathbf{E}[X] = p. \tag{3.30}$$

    $$\mathbf{VAR}[X] = \mathbf{E}[I_A^2] - \mathbf{E}[X]^2 = p - p^2 = pq. \tag{3.31}$$

- The maximum variability occurs when $p = 1/2$ which corresponds to the case that is most difficult to predict.

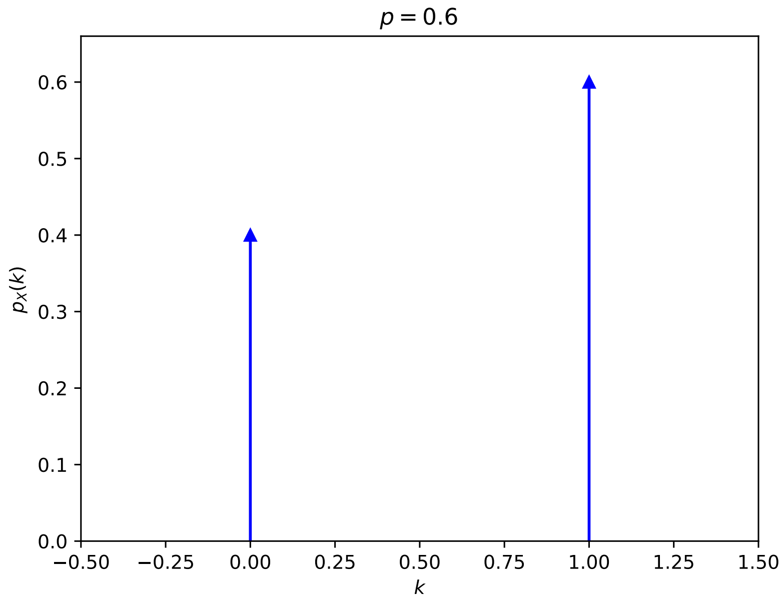

Figure 3.1: The PMF of the Bernoulli random variable with $p = 0.6$

3.5.2 The Binomial Random Variables

- The binomial random variable X is (can be) defined by the number of successes in n independent Bernoulli trials.
  - range:
  $$S_X = \{0, 1, \dots, n\}. \tag{3.32}$$
  - PMF:
  $$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{3.33}$$
  Figure 3.2 shows the PMFs of the two binomial random variables; one with with $n = 5$ and $p = 0.3$ and the other with with $n = 10$ and $p = 0.6$.
  - mean and variance:
  $$\mathbf{E}[X] = np. \tag{3.34}$$
  $$\mathbf{VAR}[X] = npq. \tag{3.35}$$

  Proof: First, let know check whether the sum of all the PMF values is 1. The binomial theorem in §A implies
  $$\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+q)^n = 1,$$

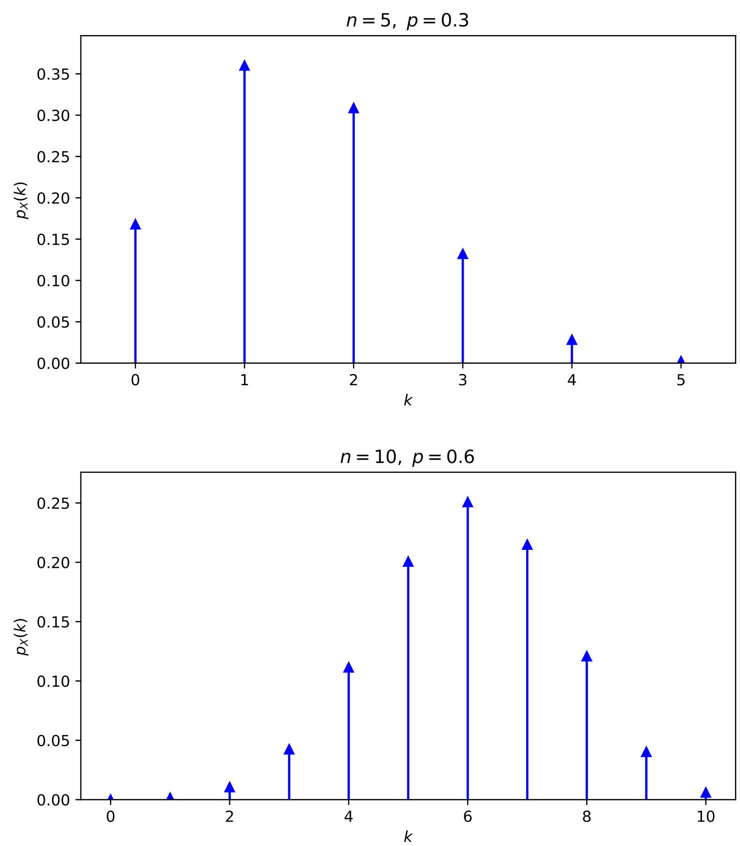

Figure 3.2: The PMFs of the two binomial random variables

hence the proof. The expected value can be evalued as

$$\begin{aligned}
\mathbf{E}[X] = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} &= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k} \\
&= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!}p^{k-1}(1-p)^{((n-1)-(k-1))} \\
&= np\sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!}p^{k-1}(1-p)^{((n-1)-k)} \\
&= np
\end{aligned}$$

where the last equality comes from that the last summation is equal to the sum of the PMFs of the binomial random variable with $n-1$ and p. The second moment of X is

$$\begin{aligned}
\mathbf{E}[X^2] = \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} &= \sum_{k=1}^{n} k\frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n-1}(k+1)\frac{n!}{(k)!(n-k-1)!}p^{k+1}(1-p)^{n-1-k} \\
&= np\sum_{k=0}^{n-1}(k+1)\frac{(n-1)!}{(k)!(n-k-1)!}p^{k}(1-p)^{n-1-k} \\
&= np\sum_{k=0}^{n-1}(k+1)\binom{n-1}{k}p^{k}(1-p)^{n-1-k} \\
&= np\sum_{k=0}^{n-1}k\binom{n-1}{k}p^{k}(1-p)^{n-1-k} + np\sum_{k=0}^{n-1}\binom{n-1}{k}p^{k}(1-p)^{n-1-k} \\
&= n(n-1)p^2 + np = n^2p^2 - np^2 + np = n^2p^2 + npq,
\end{aligned}$$

thus

$$\mathbf{VAR}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = npq.$$

The alternative proofs can be found in §B.

- The expected value $\mathbf{E}[X] = np$ **agrees with our intuition** since we expect a fraction p of the outcomes to result in success.
- The variance of the binomial is n times the variance of a Bernoulli random variable. The variance becomes maximum when $p = 1/2$ for the same reason that the maximum variance of the Bernoulli random variable is achieved when $p = 1/2$.
- The maximum PMF occurrs at $k_{\max} = \lfloor (n+1)p \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer that is smaller than or equal to x. When $(n+1)p$ is an integer, then the maximum is achieved at $k_{\max}$ and $k_{\max}+1$.
- The ratio of $p_X(k+1)$ and $p_X(k)$:

$$\frac{p_X(k+1)}{p_X(k)} = \binom{n}{k+1}p^{k+1}q^{n-k-1}\Big/\binom{n}{k}p^kq^{n-k} = \frac{n-k}{k+1}\frac{p}{q}.$$

This can be used for calculation of the PMF of the binomial random variable. Should this concern us?

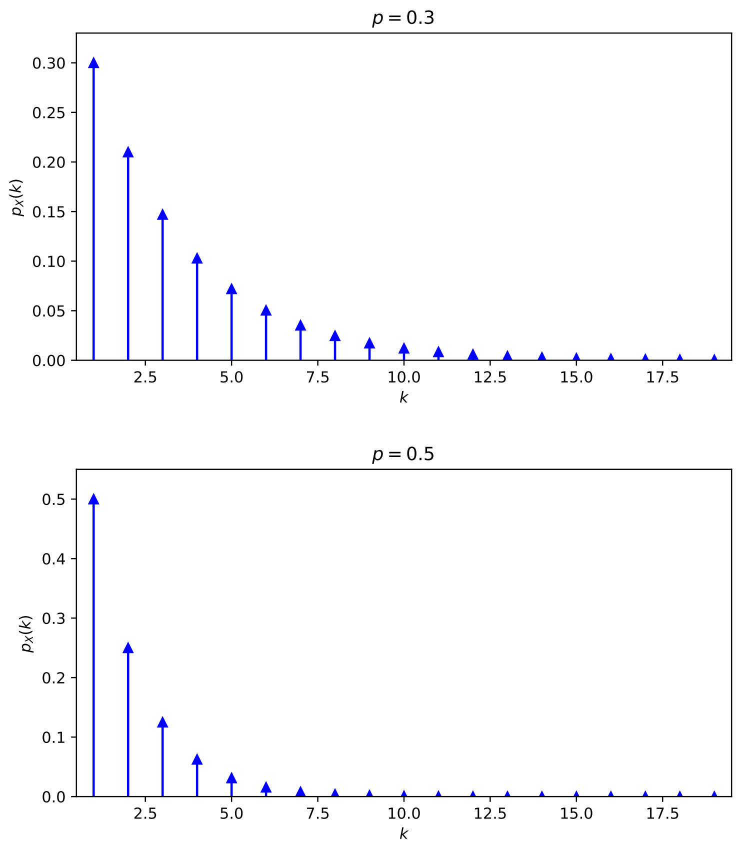

Figure 3.3: The PMFs of the two geometric random variables

3.5.3 The Geometric Random Variable

- The geometric random variable arises when we count the number of independent Bernoulli trials until the first occurrence of a success. The number X is called the **geometric random variable** and it takes on values from the set $\{1, 2, \ldots\}$.
  - range:
$$S_X = \mathbf{N} = \{1, 2, \ldots\}. \tag{3.36}$$
  - PMF:
$$p_X(k) = (1 - p)^{k-1} p. \tag{3.37}$$
    Figure 3.3 shows the PMFs of the two geometric random variables; one with $p = 0.3$ and the other with $p = 0.5$.
  - mean and variance:
$$\mathbf{E}[X] = 1/p. \tag{3.38}$$
$$\mathbf{VAR}[X] = (1 - p)/p^2. \tag{3.39}$$

- As p increases, the PMF decays more rapidly.
- **The larger p becomes, the smaller $\mathbf{E}[X]$ becomes**, which makes sense since the larger probability p, the sooner the heads comes up.
- The larger p becomes, the smaller $\mathbf{VAR}[X]$ becomes.
- The geometric random variable is **the only discrete random variable that satisfies the memoryless property**:

$$P[X \geq k + j | X > j] = P[X \geq k] \tag{3.40}$$

for all j and k. The above expression states that if a success has not occurred in the first j trials, then the probability of having to perform at least k more trials is the same as the probability of initially having to perform at least k trials. Thus, each time a failure occurs, **the system "forgets"** and begins anew as if it were performing the first trial.

3.5.4 The Poisson Random Variables

- The Poisson random variable arises in situations where the events occur **"completely at random" in time or space**. For example, the Poisson random variable arises in counts of emissions from radioactive substances, in counts of demands for telephone connections, and in counts of defects in a semiconductor chip.

Suppose that α is the average number of event occurrences in a specified time interval or region in space. Then

  - range:

$$S_X = \{0, 1, \ldots\}. \tag{3.41}$$

  - PMF:

$$p_X(k) = \frac{\alpha^k}{k!} e^{-\alpha}. \tag{3.42}$$

Figure 3.4 shows the PMFs of three Poisson random variables; $\alpha = 0.75$, $\alpha = 3$, and $\alpha = 9$.

  - mean and variance:

$$\mathbf{E}[X] = \alpha. \tag{3.43}$$

$$\mathbf{VAR}[X] = \alpha. \tag{3.44}$$

Proof: The convergence of the Taylor's series of the exponential function (A.13) implies

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

for any $x \in \mathbf{R}$, thus

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} = e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = 1.$$

Then the expected value is

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k \frac{\alpha^k}{k!} e^{-\alpha} = \alpha \sum_{k=1}^{\infty} \frac{\alpha^{k-1}}{(k-1)!} e^{-\alpha} = \alpha$$

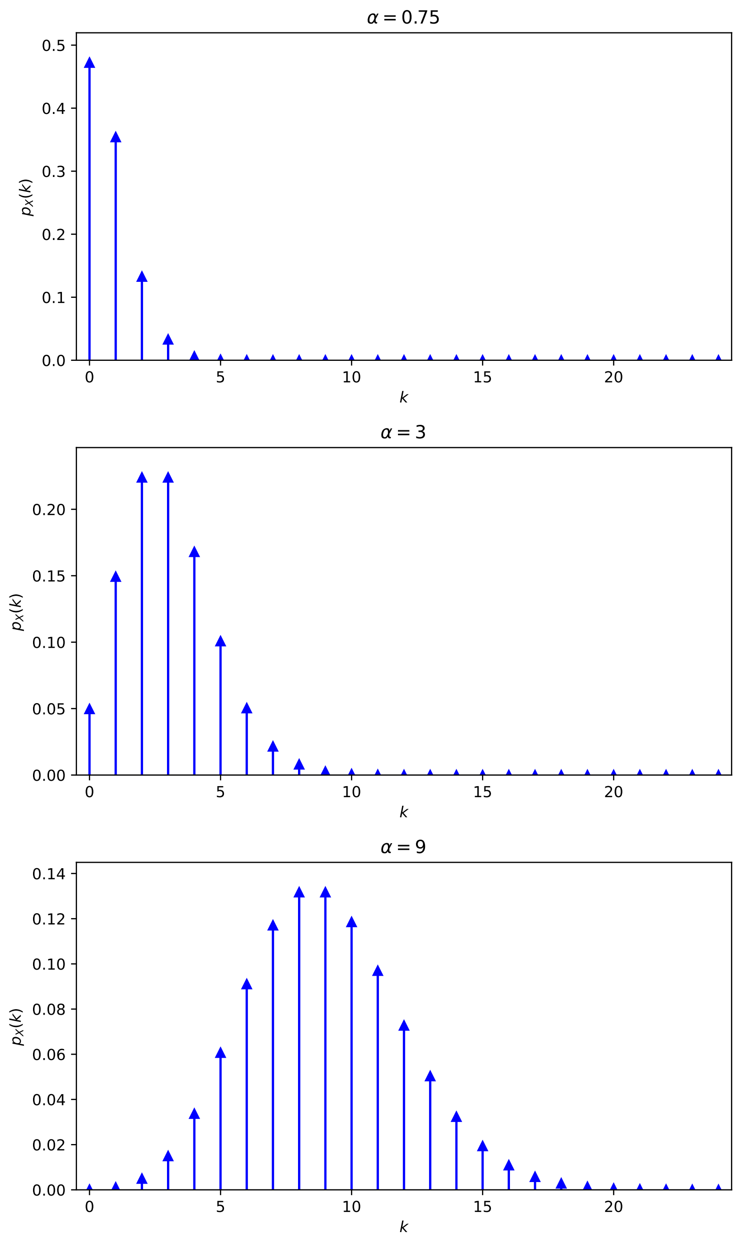

Figure 3.4: The PMFs of Poisson random variables

also by (A.13). The second moment of X is

$$\begin{aligned}
\mathbf{E}[X^2] = \sum_{k=1}^{\infty} k^2 p_X(k) = \sum_{k=1}^{\infty} k \frac{\alpha^k}{(k-1)!} e^{-\alpha} = \sum_{k=1}^{\infty} ((k-1)+1) \frac{\alpha^k}{(k-1)!} e^{-\alpha} \\
&= \sum_{k=2}^{\infty} \frac{\alpha^k}{(k-2)!} e^{-\alpha} + \sum_{k=1}^{\infty} \frac{\alpha^k}{(k-1)!} e^{-\alpha} \\
&= \alpha^2 \sum_{k=2}^{\infty} \frac{\alpha^{k-2}}{(k-2)!} e^{-\alpha} + \alpha \sum_{k=1}^{\infty} \frac{\alpha^{k-1}}{(k-1)!} e^{-\alpha} \\
&= \alpha^2 + \alpha,
\end{aligned}$$

hence

$$\mathbf{VAR}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \alpha.$$

∎

- For $\alpha < 1$, $p_X(k)$ is maximum at $k = 0$; for $\alpha > 1$, $p_X(k)$ is maximum at $\lfloor \alpha \rfloor$; if α is a positive integer, the $p_X(k)$ is maximum at $k = \alpha$ and at $k = \alpha - 1$.

- LGExample 3.30 **Queries at a Call Center** The number N of queries arriving in t seconds at a call center is a Poisson random variable with $\alpha = \lambda t$ where λ is the average arrival rate in queries/second. Assume that the arrival rate is four queries per minute. Find the probability of the following events: (a) more than 4 queries in 10 seconds; (b) fewer than 5 queries in 2 minutes.

  In part (a), the average number of queries for 10 seconds is $\alpha = 4/6 = 2/3$. Thus, the probability that more than 4 queries arrive in 10 seconds is

  $$P[X > 4] = 1 - P[X \leq 4] = 1 - \sum_{k=0}^{4} \frac{(2/3)^k}{k!} e^{-(2/3)} = 6.33 \times 10^{-4}.$$

  In part (b), the average number of queries for 2 minutes is $\alpha = 4 \cdot 2 = 8$. Thus, the probability that fewer than 5 queries arrive in 2 minutes is

  $$P[X < 5] = \sum_{k=0}^{4} \frac{8^k}{k!} e^{-8} = 0.996.$$

  These values were obtained by the following python code.

```
>>> import scipy.stats as ss
>>> rv = ss.poisson(2/3.)
>>> 1-rv.cdf(4)
0.00063252138923441947
>>> rv = ss.poisson(8.)
>>> rv.cdf(4)
0.099632400487045969
>>>
```

- One of the applications of the Poisson probabilities in (3.42) is to **approximate the binomial probabilities in the case where** p **is very small and** n **is very large**, that is, where the event A of interest is very rare but the number of Bernoulli trials is very large. If $\alpha = np$ is fixed, then as n becomes large:

  $$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\alpha^k}{k!} e^{-\alpha} \tag{3.45}$$

  for $k = 0, 1, \ldots$.

Proof: If we let $\alpha = np$,

$$\begin{aligned}\binom{n}{k} p^k (1-p)^{n-k} &= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\ &= \frac{(1-p)^{-k}}{k!} n(n-1)\cdots(n-k+1) p^k (1-p)^n \\ &= \frac{(1-p)^{-k}}{k!} (np)(np-p)\cdots(np-(k-1)p)(1-\alpha/n)^n \\ &= \frac{(1-p)^{-k}}{k!} (\alpha)(\alpha-p)\cdots(\alpha-(k-1)p)(1-\alpha/n)^n.\end{aligned}$$

Now if we fix $\alpha = np$ and let n go to ∞, p goes to 0, hence

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\alpha^k}{k!} e^{-\alpha}$$

where we use

$$\lim_{n\to\infty} (1-\alpha/n)^n = \lim_{n\to\infty} (1+1/n)^{-\alpha n} = \left(\lim_{n\to\infty} (1+1/n)^n\right)^{-\alpha} = e^{-\alpha}.$$

- The Poisson random variable appears in numerous physical situations because many models are very large in scale and involve very rare events.
  - For example, the Poisson PMF gives an accurate prediction for the relative frequencies of the number of particles emitted by a radioactive mass during a fixed time period. This correspondence can be explained as follows. A radioactive mass is composed of a large number of atoms, say n. In a fixed time interval each atom has a very small probability p of disintegrating and emitting a radioactive particle. If atoms disintegrate independently of other atoms, then the number of emissions in a time interval can be viewed as the number of successes in n trials. For example, one microgram of radium contains about $n = 10^{16}$ atoms, and the probability that a single atom will disintegrate during a one-millisecond time interval is $p = 10^{-15}$ [Rozanov, p. 58]. Thus it is an understatement to say that the conditions for the approximation in (3.45) hold: n is so large and p so small that one could argue that the limit $n \to \infty$ has been carried out and that the number of emissions is **exactly** a Poisson random variable.

3.5.5 The Uniform Random Variable

- The discrete uniform random variable X takes on values in a set of consecutive integers $S_X = \{j+1, \ldots, j+L\}$ with equal probability:
  - range:
  $$S_X = \{j+1, j+2, \ldots, j+L\}. \tag{3.46}$$
  - PMF:
  $$p_X(k) = \frac{1}{L}. \tag{3.47}$$
  - mean and variance:
  $$\mathbf{E}[X] = j + \frac{L+1}{2}. \tag{3.48}$$
  $$\mathbf{VAR}[X] = \frac{L^2-1}{12}. \tag{3.49}$$
- This humble random variable occurs whenever outcomes are equally likely, *e.g.*, toss of a fair coin or a fair die, spinning of an arrow in a wheel divided into equal segments, selection of numbers from an urn.

3.5.6 The Zipf Random Variable

- The Zipf random variable is named for George Zipf who observed that the frequency of words in a large body of text is proportional to their rank. Suppose that words are ranked from most frequent, to next most frequent, and so on. Let X be the rank of a word, then $S_X = \{1, 2, ..., L\}$ where L is the number of distinct words.

  – range:
  $$S_X = \{1, 2, \ldots, L\}. \tag{3.50}$$

  – PMF:
  $$p_X(k) = \frac{1}{c_L}\frac{1}{k}. \tag{3.51}$$

  where c_L is a normalization constant:
  $$c_L = \sum_{j=1}^{L} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{L}. \tag{3.52}$$

  – mean and variance:
  $$\mathbf{E}[X] = \frac{L}{c_L}. \tag{3.53}$$
  $$\mathbf{VAR}[X] = \frac{L(L+1)}{2c_L} - \frac{L^2}{c_L^2}. \tag{3.54}$$

- The constant c_L occurs frequently in calculus and is called the Lth harmonic mean and increases approximately as $\log L$. For example, for $L = 100$, $c_L = 5.187378$ and $c_L - \log L = 0.582207$. It can be shown that as $L \to \infty$, $c_L - \log L \to 0.57721$.

- The Zipf and related random variables have gained prominence with the growth of the Internet where they have been found in a variety of measurement studies involving Web page sizes, Web access behavior, and Web page interconnectivity. These random variables had previously been found extensively in studies on the distribution of wealth and, not surprisingly, are now found in Internet video rentals and book sales.

3.6 Generation of Discrete Random Variables

MAYBE LATER WITH OTHER RELATED TOPICS

Chapter 4

One Random Variable

4.1 The Cumulative Distribution Function

- The **cumulative distribution function** has the advantage that it is **not limited** to discrete random variables and applies to all types of random variables.
- **Definition 4.1** *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 $\mathbf{R}$ with the property that the set $A_b = \{\zeta | X(\zeta) \leq b\}$ is in $\mathcal{F}$ for every b in $\mathbf{R}$.*
- The **cumulative distribution function** (CDF) of a random variable X is defined as the probability of the event $\{X \leq x\}$:
$$F_X(x) = P[X \leq x] \tag{4.1}$$
for $x \in \mathbf{R}$.
  - It is the probability that the random variable X takes on a value in the set $(-\infty, x]$. In terms of the underlying sample space, the CDF is the probability of the event $\{\zeta | X(\zeta) \leq x\}$.
  - $F_X(x)$ **is a function of the variable x.**
- The events of interest when dealing with numbers are intervals of the real line, and their complements, unions, and intersections. We show below that the probabilities of all of these events can be expressed in terms of the CDF.
- LGExample 4.1 **Three Coin Tosses** Figure 4.1 shows the CDF and probability density function (PDF) of the binomial random variable with $n = 3$ and $p = 1/2$.
  - The CDF is continuous from the right and equal to 1/2 at the point $x = 1$.
$$\lim_{x \to 1+0} F_X(x) = 1/2, \quad \lim_{x \to 1-0} F_X(x) = 1/8.$$
  - Indeed, we note the magnitude of the jump at the point $x = 1$ is equal to $P[X = 1] = 1/2 - 1/8 = 3/8$.
- LGExample 4.2 **Uniform Random Variable in the Unit Interval** Figure 4.2 shows the CDF and PDF of the uniform random variable with support $[0, 1]$.
  - No discontinuity in the CDF
  - The derivative of CDF coincides with the PDF (except two points, $x = 0$ and $x = 1$).

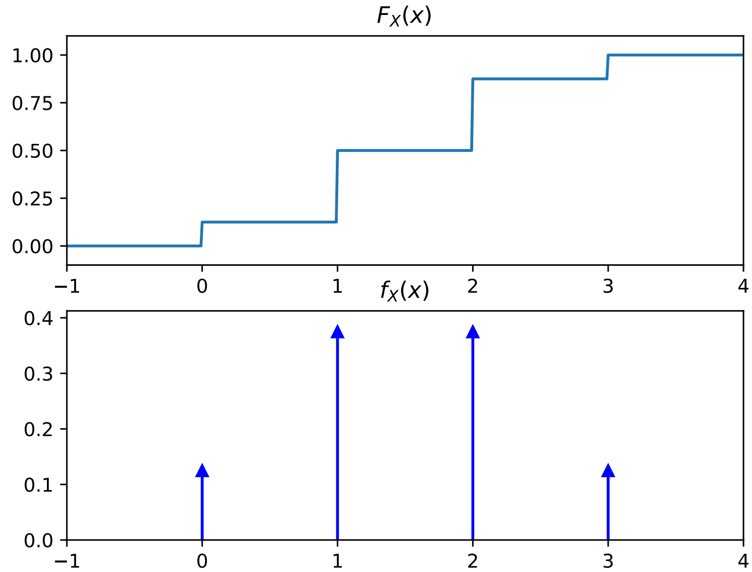

Figure 4.1: The CDF and PDF of the binomial random variable with $n = 3$ and $p = 1/2$.

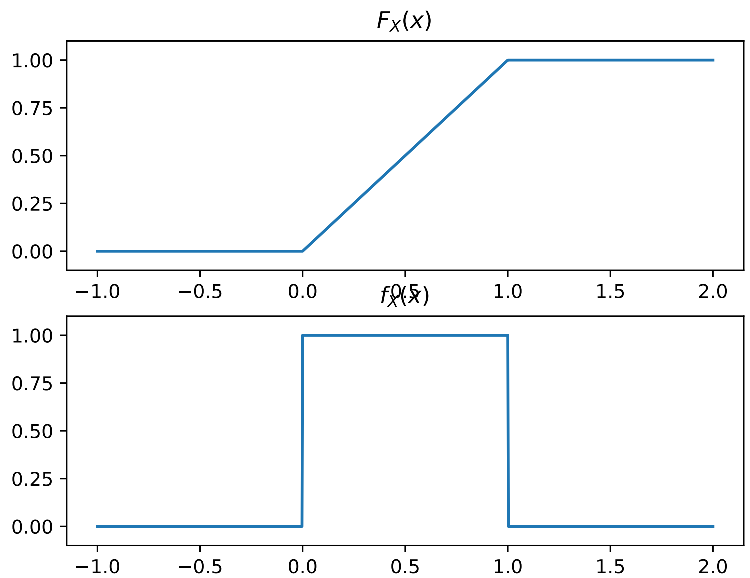

Figure 4.2: The CDF and PDF of the uniform random variable with support $[0, 1]$.

- Basic properties of the CDF: The axioms of probability and their corollaries imply that the CDF has the following properties:

  (i) $0 \leq F_X(x) \leq 1$.

  (ii) $\lim_{x\to\infty} F_X(x) = 1$.

  (iii) $\lim_{x\to-\infty} F_X(x) = 0$.

  (iv) $F_X(x)$ is a nondecreasing function of x, that is, if $a < b$, then $F_X(a) \leq F_X(b)$.

  (v) $F_X(x)$ is continuous from the right, *i.e.*,

  $$F_X(x) = \lim_{h\to+0} F_X(x+h) = F_X(x^+).$$

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

  (vii) $P[X = b] = F_X(b) - F_X(b^-)$.

  (viii) $P[X > x] = 1 - F_X(x)$.

  - The first five properties confirm that, in general, the CDF is a nondecreasing function that grows from 0 to 1 as x increases from $-\infty$ to ∞.
  - Property (vii) states that the probability that $X = b$ is given by the magnitude of the jump of the cdf at the point b. This implies that if the CDF is continuous at a point b, then $P[X = b] = 0$.
  - Properties (vi) and (vii) can be combined to compute the probabilities of other types of intervals. For example, since $\{a \leq X \leq b\} = \{X = a\} \cup \{a < X \leq b\}$, then

    $$P[a \leq X \leq b] = P[X = a] + P[a < X \leq b] = F_X(a) - F_X(a^-) + F_X(b) - F_X(a) = F_X(b) - F_X(a^-).$$

  - While intuitively clear, **properties (ii), (iii), (v), and (vii) require more advanced limiting arguments.**

4.1.1 The Three Types of Random Variables

- A **discrete random variables** has a CDF that is a right-continuous, staircase function of x, with jumps at a countable set of points. The CDF $F_X(x)$ of a discrete random variable is the sum of the probabilities of the outcomes less than x and **can be written as the weighted sum of unit step functions**

  $$F_X(x) = \sum_{x_k \leq x} p_X(x_k) = \sum_k p_X(x_k) u(x - x_k) \tag{4.2}$$

  where the PMF $p_X(x_k)$ gives the magnitude of the jumps in the CDF.

- A **continuous random variable** is defined as a random variable whose **CDF** $F_X(x)$ **is continuous everywhere, and which, in addition, is sufficiently smooth** that it can be written as an **integral of some nonnegative function** $f(x)$:

  $$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx. \tag{4.3}$$

  - The continuity of the CDF and property (vii) implies that continuous random variables have $P[X = x] = 0$ **for all** x. **Every possible outcome has probability zero!**
  - We calculate probabilities as integrals of "probability densities" over intervals of the real line, *i.e.*,

    $$P[a \leq X \leq b] = \int_a^b f_X(x).$$

- A **random variable of mixed type** is a random variable with a CDF that has jumps on a countable set of points, but that also increases continuously over **at least one interval** of values of x.
  - The CDF for these random variables has the form

$$F_X(x) = pF_{X_1}(x) + (1-p)F_{X_2}(x)$$

    where $0 < p < 1$, and $F_{X_1}(x)$ is the CDF of a discrete random variable, X_1, and $F_{X_2}(x)$ is the CDF of a continuous random variable, X_2.

4.1.2 Fine Point: Limiting properties of CDF †

- Properties (ii), (iii), (v), and (vii) require the continuity property of the probability function discussed in §2.9.
- For property (ii), we consider the sequence of events $\{X \le n\}$ which increases to include all of the sample space S as n approaches ∞, that is, all outcomes lead to a value of X less than infinity. The continuity property of the probability function (Corollary 2.8) implies that:

$$\lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} P[X \le n] = P\left[\lim_{n\to\infty} \{X \le n\}\right] = P[S] = 1.$$

- For property (iii), we take the sequence $\{X \le n\}$ which decreases to the empty set $\emptyset$, that is, no outcome leads to a value of X less than $-\infty$:

$$\lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} P[X \le -n] = P\left[\lim_{n\to\infty} \{X \le -n\}\right] = P[\emptyset] = 0.$$

- For property (v), we take the sequence of events $\{X \le b + 1/n\}$ which decreases to $\{X \le b\}$ from the right:

$$\begin{aligned} \lim_{n\to\infty} F_X(b+1/n) &= \lim_{n\to\infty} P[X \le b + 1/n] \\ &= P\left[\lim_{n\to\infty} \{X \le b + 1/n\}\right] = P[\{X \le b\}] = F_X(b). \end{aligned}$$

- For property (vii), we take the sequence of events, $\{b - 1/n < X \le b\}$ which decreases to $\{b\}$ from the left:

$$\begin{aligned} \lim_{n\to\infty} (F_X(b) - F_X(b-1/n)) &= \lim_{n\to\infty} P[b - 1/n < X \le b] \\ &= P\left[\lim_{n\to\infty} \{b - 1/n < X \le b\}\right] = P[X = b]. \end{aligned}$$

4.2 The Probability Density Function

- The probability density function (PDF) of X, if it exists, is defined as the derivative of $F_X(x)$:

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{4.4}$$

- The PDF is an alternative, and more useful, way of specifying the information contained in the cumulative distribution function.
- The PDF represents the "density" of probability at the point x in the following sense: The probability that X is in a small interval in the vicinity of x, *i.e.*,

$$P[x < X \le x + h] = F_X(x+h) - F_X(x) = \frac{F_X(x+h) - F_X(x)}{h} h \simeq f_X(x)h$$

  for very small h.

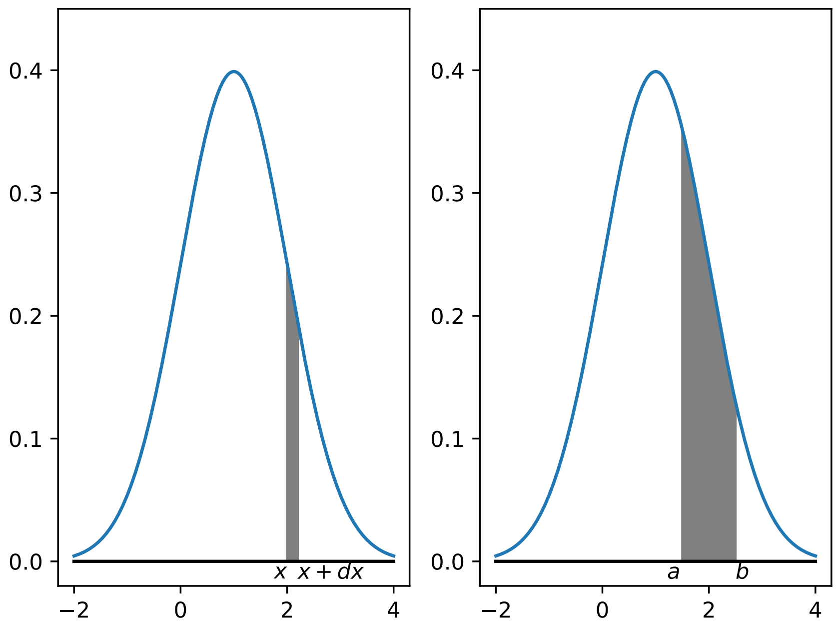

Figure 4.3: (left) The PDF specifies the probability of intervals of infinitesimal width. (right) The probability of an interval $[a, b]$ is the area under the PDF in the interval.

- Properties of the PMF:
  (i) Since $F_X(x)$ is nondecreasing,
  $$f_X(x) \geq 0 \tag{4.5}$$
  (ii) (Almost) by definition,
  $$P[a \leq X \leq b] = \int_a^b f_X(x)\,dx. \tag{4.6}$$
  – The probability of an interval is therefore the area under $f_X(x)$ in that interval, as shown in Figure 4.3.

  (iii) The CDF of X can be obtained by integrating the PDF:
  $$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt. \tag{4.7}$$
  – Thus the PDF completely specifies the behavior of continuous random variables.

  (iv) By letting x tend to infinity in (4.7), we obtain a **normalization** condition for the PDF:
  $$1 = \int_{-\infty}^{\infty} f_X(x)\,dx. \tag{4.8}$$
- A valid PDF can be formed from any nonnegative, piecewise continuous function $g(x)$ that has a finite integral:
  $$\int_{-\infty}^{\infty} g(x)\,dx = c < \infty. \tag{4.9}$$

- LGExample 4.6 **Uniform Random Variable** The PDF of the uniform random variable is given by

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

and the CDF is given by

$$F_X(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \le x \le b \\ 1 & x > b \end{cases}$$

as in Figure 4.2.

- LGExample 4.7 **Exponential Random Variable** The PDF and CDF are given by

$$f_X(x) = \lambda e^{-\lambda x} u(x)$$

and

$$f_X(x) = (1 - e^{-\lambda x}) u(x)$$

where $u(x)$ is the unit step function:

$$u(x) = \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

4.2.1 PDF of Discrete Random Variable

- The derivative of the CDF does **not** exist at points where the CDF is **not continuous**. Thus the notion of PDF as defined by (4.4) does not apply to discrete random variables at the points where the CDF is discontinuous.
- We can generalize the definition of the probability density function by noting the relation between the **unit step function** and the **delta function**.
- The **unit step function** is defined as

$$u(x) = \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{4.10}$$

- The **(Dirac) delta function** $\delta(x)$ is related to the unit step function bu the following equation:

$$u(x) = \int_{-\infty}^{x} \delta(t)\, dt \tag{4.11}$$

- The translated unit step function is then:

$$u(x - x_0) = \int_{-\infty}^{x - x_0} \delta(t)\, dt = \int_{-\infty}^{x} \delta(t - x_0)\, dt \tag{4.12}$$

- Substituting (4.12) into (4.2) yields

$$F_X(x) = \sum_k p_X(x_k) u(x - x_k) = \sum_k p_X(x_k) \int_{-\infty}^{x} \delta(t - x_k)\, dt = \int_{-\infty}^{x} \sum_k p_X(x_k) \delta(t - x_k)\, dt$$

- This suggests that we define the PDF for a discrete random variable by

$$f_X(x) = \frac{d}{dx} F_X(x) = \sum_k p_X(x_k) \delta(t - x_k)\, dt \tag{4.13}$$

– Thus the generalized definition of PDF places a delta function of weight $P[X = x_k]$ at the points x_k where the CDF is discontinuous.

- **Please read the parts explaining the delta function**!
  - The Dirac delta function has the following property:

$$\int_{-\infty}^{\infty} g(t)\delta(t-x)\,dt = g(x). \tag{4.14}$$

- The **PDF of a random variable of mixed type** will also contain delta functions at the points where its CDF is not continuous.
- LGExample 4.9 The binomial random variable with $n = 3$ and $p = 1/2$. The CDF is

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x-1) + \frac{3}{8}u(x-2) + \frac{1}{8}u(x-3)$$

and the PDF is

$$f_X(x) = \frac{1}{8}\delta(x) + \frac{3}{8}\delta(x-1) + \frac{3}{8}\delta(x-2) + \frac{1}{8}\delta(x-3).$$

4.2.2 Conditional CDFs and PDFs

- The **conditional CDF** of X given the event C is defined by

$$F_X(x|C) = \frac{P[\{X \leq x\}|\cap]C}{P[C]} \tag{4.15}$$

where $P[C] > 0$.

  - easy to show that $F_X(x|C)$ satisfies all the properties of a CDF. (See Problem 4.29.) The conditional pdf of X given C is then defined by

- The **conditional PDF** of X given the event C is defined by

$$f_X(x|C) = \frac{d}{dx}F_X(x|C) \tag{4.16}$$

- Suppose that we have a partition of the sample space S into the union of disjoint events B_1, B_2, ..., B_n. Let $F_X(x|B_i)$ be the conditional CDF of X given event B_i. The **theorem on total probability** allows us to find the CDF of X in terms of the conditional CDFs:

$$F_X(x|\,=)P[X \leq x] = \sum_{i=1}^{n} P[X \leq x|B_i]P[B_i] = \sum_{i=1}^{n} F_X(x|B_i)P[B_i] \tag{4.17}$$

  - The PDF is obtained by differentiation:

$$f_X(x) = \frac{d}{dx}F_X(x) = \sum_{i=1}^{n} f_X(x|B_i)P[B_i]. \tag{4.18}$$

- LGExample 4.11 **A binary transmission system** A binary transmission system sends a "0" bit by transmitting a $-v$ voltage signal, and a "1" bit by transmitting a $+v$. The received signal is corrupted by Gaussian noise and given by:

$$Y = X + N$$

where X is the transmitted signal, and N is a noise voltage with PDF $f_N(x)$. Assume that $P[``1"] = p = 1 = P[``0"]$ and that N is uniform random variable with 0 and σ as its mean and standard deviation, *i.e.*,

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

Find the PDF of Y.

– The PDF is

$$f_Y(x) = f_Y(x|B_0)P[B_0] + f_Y(x|B_1)P[B_1] = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x+v)^2/2\sigma^2}(1-p) + \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-v)^2/2\sigma^2} p.$$

– LGFig 4.5 shows the two conditional PDFs. The transmitted signal X shifts the center of mass of the Gaussian PDF.

4.3 The Expected Value of X

- The **expected value** or **mean** of a random variable X is defined by

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx. \tag{4.19}$$

  The expected value $\mathbf{E}[X]$ is defined if the above integral converges absolutely, that is,

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} |x| f_X(x)\, dx. \tag{4.20}$$

- If we view $f_X(x)$ as the **distribution of mass on the real line**, then $\mathbf{E}[X]$ represents the **center of mass** of this distribution.

- The definition in (4.19) is applicable if we express the PDF of a discrete random variable using delta functions:

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x \sum_k p_X(x_k)\delta(x - x_k)\, dx = \sum_k p_X(x_k) \int_{-\infty}^{\infty} x\delta(x - x_k)\, dx$$

  where (4.14) is used for the last equality.

- LGExample 4.12 **Mean of a Uniform Random Variable**

  The PDF of the uniform random variable is

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise,} \end{cases}$$

  hence

$$\mathbf{E}[X] = \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a} \cdot \frac{1}{2}(b-a)^2 = \frac{a+b}{2}.$$

- If $f_X(m - x) = f_X(m + x)$ for all $x \in \mathbf{R}$, then $\mathbf{E}[X] = m$ since

$$0 = \int_{-\infty}^{\infty} (m-x) f_X(x)\, dx = \int_{-\infty}^{\infty} m f_X(x)\, dx - \int_{-\infty}^{\infty} x f_X(x)\, dx = m - \mathbf{E}[X].$$

- LGExample 4.13 **Mean of a Gaussian Random Variable**

  The PDF of a Gaussian random variable $e^{-(x-m)^2/\sigma^2}/\sqrt{2\pi}\sigma$ is symmetric about the point $x = m$, hence $\mathbf{E}[X] = m$.

- Aother formulas for $\mathbf{E}[X]$:
  - For nonnegative continuous random variables:

$$\mathbf{E}[X] = \int_0^{\infty} (1 - F_X(x))\, dx \tag{4.21}$$

  - For nonnegative integer-valued random variables:

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} P[X > k] \tag{4.22}$$

- LGExample 4.14 **Mean of Exponential Random Variable**

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} \lambda e^{-\lambda x}\, dx$$

4.4 Important Continuous Random Variables

4.5 Functions of a Random Variable

4.6 The Markov and Chebyshev Inequalities

4.7 Transform Methods

4.8 Basic Reliability Calculations

4.9 Computer Methods for Generating Random Variables

4.10 Entropy †

Appendix A

Mathematics

A.1 The binomial theorem

- The binomial theorem: for any a, b, and $n \in \mathbf{N}$,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \tag{A.1}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!} \tag{A.2}$$

is the number of combinations of choosing k items out of n items, which is read as "n choose k" or sometimes "n combination k". It is sometimes denoted by ${}_nC_k$. (Note the convension defined $0! = 1$.)

Proof: It is obvious that (A.1) holds for $n = 1$. Suppose that (A.1) holds for $n = m$. Then the inductive assumption implies

$$\begin{aligned}
(a+b)^{m+1} &= (a+b)(a+b)^m = (a+b)\sum_{k=0}^{m}\binom{m}{k}a^k b^{m-k} \\
&= \sum_{k=0}^{m}\binom{m}{k}a^k b^{m-k+1} + \sum_{k=0}^{m}\binom{m}{k}a^{k+1}b^{m-k} \\
&= \sum_{k=0}^{m}\binom{m}{k}a^k b^{m-k+1} + \sum_{k'=1}^{m+1}\binom{m}{k'-1}a^k b^{m-k'+1} \\
&= \sum_{k=0}^{m}\binom{m}{k}a^k b^{m-k+1} + \sum_{k'=1}^{m+1}\binom{m}{k'-1}a^{k'} b^{m-k'+1} \\
&= a^0 b^{m+1} + \sum_{k=1}^{m}\left(\binom{m}{k} + \binom{m}{k-1}\right)a^k b^{m+1-k} + a^{m+1}b^0 \\
&= a^0 b^{m+1} + \sum_{k=1}^{m}\binom{m+1}{k}a^k b^{m+1-k} + a^{m+1}b^0 \\
&= \sum_{k=0}^{m+1}\binom{m+1}{k}a^k b^{m+1-k}
\end{aligned}$$

where the following formula is used for the last equality:

$$\begin{aligned}\binom{m}{k}+\binom{m}{k-1} &= \frac{m!}{k!(m-k)!}+\frac{m!}{(k-1)!(m-k+1)!}\\ &= \frac{m!(m-k+1)+m!k}{k!(m-k+1)!}=\frac{(m+1)!}{k!(m-k+1)!}=\binom{m+1}{k}\end{aligned}$$

Thus, (A.1) holds for $n = m+1$, too. Therefore the mathematical induction implies (A.1) holds for all $n \in \mathbf{N}$. ∎

- If we differentiate both sides of (A.1) with respect to a and multiplying both sides by a, we have

$$\sum_{k=1}^{n}\binom{n}{k}ka^k b^{n-k} = na(a+b)^{n-1}. \tag{A.3}$$

If we again differentiate both sides of (A.3) with respect to a and multiplying both sides by a, we have

$$\sum_{k=0}^{n}\binom{n}{k}k^2 a^{k-1} b^{n-k} = na(a+b)^{n-2}(a(n-1)+a+b). \tag{A.4}$$

A.2 Inifinite sequences and series

All the following equations hold only for $x \in (-1,1)$.

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \tag{A.5}$$

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}, \tag{A.6}$$

$$\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}, \tag{A.7}$$

$$\sum_{k=3}^{\infty} k(k-1)(k-2)x^{k-3} = \frac{6}{(1-x)^4}, \tag{A.8}$$

$$\sum_{k=m}^{\infty} k(k-1)\cdots(k-m+1)x^{k-m} = \frac{m!}{(1-x)^{m+1}}. \tag{A.9}$$

Multiplying x on both sides of (A.6) yields

$$\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2}. \tag{A.10}$$

Multiplying x^2 on both sides of (A.7) and adding it to (A.10) yields

$$\sum_{k=1}^{\infty} k^2x^k = \frac{2x^2}{(1-x)^3}+\frac{x}{(1-x)^2}. = \frac{2x^2+x(1-x)}{(1-x)^3} = \frac{x^2+x}{(1-x)^3} = \frac{x(1+x)}{(1-x)^3} \tag{A.11}$$

A.3 Taylor series

- The Taylor series of a real or complex function $f(x)$ that is infinitely differentiable in a neighborhood of a real or complex number a is the power series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots \tag{A.12}$$

  where $n!$ denotes the factorial of n and $f^{(n)}(a)$ denotes the nth derivative of f evaluated at the point a. The zeroth derivative of f is defined to be f itself and $(x-a)^0$ and 0! are both defined to be 1. In the case that a = 0, the series is also called a **Maclaurin series**.

- A function is **analytic** in an open disc centered at a if and only if its Taylor series converges to the value of the function at each point of the disc.

- If $f(x)$ is equal to its Taylor series everywhere it is called **entire**.
  - The polynomials and the exponential function e^x and the trigonometric functions (sine and cosine) are examples of entire functions.
  - Examples of functions that are not entire include the logarithm, the trigonometric function tangent, and its inverse arctan. For these functions the Taylor series do not converge if x is far from a.
  - Taylor series can be used to calculate the value of an entire function in every point, if the value of the function, and of all of its derivatives, are known at a single point.

- The exponential function is entire, thus

$$\exp(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n \tag{A.13}$$

  for all $x \in \mathbf{C}$.

Appendix B

Proofs

- Alternative proofs for the expected value and the variance of binomial random variable in §3.5.2

  1. By regarding the binomial random variable as a sum of n independent and identically distributed Bernoulli random variables:

     Proof: Let X be the binomial random variable with n and p. If we let X_i for $i = 1, 2, \ldots, n$ be the n independent and identically distributed Bernoulli random variables with p,

     $$X = \sum_{k=1}^{n} X_i.$$

     Then the linearity of the expected value operator implies

     $$\mathbf{E}[X] = \sum_{k=1}^{n} \mathbf{E}[X_i] = np$$

     and the independence of X_i implies

     $$\mathbf{E}[X^2] = \sum_{k=1}^{n} \mathbf{E}[X_i^2] + 2\sum_{i<j} \mathbf{E}[X_i]\mathbf{E}[X_j] = n(p(1-p)+p^2)+n(n-1)p^2 = np(1-p)+(np)^2,$$

     thus

     $$\mathbf{VAR}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = np(1-p),$$

     hence the proof.

  2. Using (A.3) and (A.4):

     Proof: Let X be the binomial random variable with n and p. Then (A.3) implies

     $$\mathbf{E}[X] = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = np(p + (1-p))^{n-1} = np$$

     and (A.4) implies

     $$\mathbf{E}[X^2] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = np(p(n-1)+1) = (np)^2 + np(1-p),$$

     thus

     $$\mathbf{VAR}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = np(1-p),$$

     hence the proof.