Chapter 1 Probability and Distributions

1.1 Introduction

Question: What is probability?

Answer:

- **Probability** is the chance for something to happen, so it's always a number in $[0, 1]$.
- In addition, in this class, probability is also the name of the mathematical tool we are going to use.

How do we learn probability?

- Mathematics is the language we use to communicate with the universe. Its grammars are rules called **axioms and theorems**.
- When we learn probability, we learn those rules. So prepare for a bit of a headache at the beginning similar to learning a second language.

Relationship between probability and statistics?

- **Probability**: From population to sample.
- **Statistics**: From sample to population.

Random experiment

- The experiment can be repeated under the same condition.
- Each experiment terminates with an **outcome**.
- The outcome cannot be predicted with certainty prior to the performance of the experiment.
- The collection of all possible outcomes can be described prior to the performance of the experiment.

This collection is called the **sample space**, usually denoted by $\mathcal{C}$.

Example (1.1.1)

1. Toss a coin once, sample space $C = \{T, H\}$.
2. Toss a coin twice, sample space?

Event

A subset of the sample space $\mathcal{C}$, usually denoted by C, is called **an event**. If an outcome belongs to C, then we say that the event C has occurred.

Example (1.1.2)

1. Casting one red and one white die. Sample space is 36 ordered pairs,

$$\mathcal{C} = \{(1,1), \ldots, (1,6), (2,1), \ldots, (2,6), \ldots, (6,6)\}.$$

2. Event: sum of numbers on dice is 9,

$$C = \{(3,6), (4.5), (5,4), (6,3)\}$$

Probability

Probability quantifies the notion of chance or likelihood of an event.
Relative frequency is **an empirical definition of probability**:

- Suppose the experiment is repeated N times.
- Let k_N denote the number of times the event C actually occurred.
- The $f_N = k_N/N$ is the relative frequency of the event C in the repeated experiments.

Probability: frequentist

- Relative frequency: $f_N = k_N/N$.
- Suppose that N increases.
- Suppose that $p = \lim_{N\to\infty} f_N$ exists. Note $p \in [0, 1]$.
- Then p is the **probability** of the event C.

Example (1.1.2, cont'd)

Sum of numbers on dice is 9:

$$C = \{(3, 6), (4.5), (5, 4), (6, 3)\}$$

If each of the 36 outcomes is equally likely, then $p = 4/36$.

- We denote the probability of an event C by $P(C)$.
- It is the long-run relative frequency of the event C in a very large number of independent replications of the experiment.

Subjective probability

Consider the event $C =$ Hawkeye wins NCAA basketball championship in 2019. Suppose that I offer you two lottery tickets, and you can choose between them.

- If you pick lottery ticket 1, then we spin a roulette wheel that has 100 slots and has been declared "fair" by the Nevada Gaming Commission. If the ball lands in a slot from 1 through 10 you get $\$100$ and otherwise you get $\$0$.
- If you pick lottery ticket 2, then you get $\$100$ if C occurs and otherwise you get $\$0$.

If you choose Lottery ticket 1, then your subjective $p \leq 0.10$, and if you choose Lottery ticket 2 then your subjective $p \geq 0.10$.

- The advantage of subjective probability is that it can be extended to experiments that cannot be repeated. (Think about betting in sports, investing your money, ...)
- Mathematically, the two concepts of probability are identical. In many situations we will not need to make the distinction.

Chapter 1 Probability and Distributions

1.2 Set Theory

Sets

- A **set** is a collection of objects.
  If an element x belongs to a set C, then we write $x \in C$.
- If each element of a set C_1 is also an element of another set C_2, then C_1 is called a **subset** of C_2, written as $C_1 \subset C_2$.
- If $C_1 \subset C_2$ and $C_2 \subset C_1$, then $C_1 = C_2$.

Example (1.2.1)

Define sets $C_1 = \{x : 0 \leq x \leq 1\}$ and $C_2 = \{x : -1 \leq x \leq 2\}$.
We have $C_1 \subset C_2$.

Example (1.2.2)

Define sets $C_1 = \{(x, y) : 0 \leq x = y \leq 1\}$ and
$C_2 = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. We have $C_1 \subset C_2$.

- If a set C has no elements, then C is called the **null set**, written as $C = \phi$.

Union

- The set of all elements that belong to at least one of the sets C_1 and C_2 is called the **union** of C_1 and C_2, written as $C_1 \cup C_2$.
- For example, if $C_1 = \{1, 2, 3\}$, $C_2 = \{2, 3, 5\}$, then $C_1 \cup C_2 = \{1, 2, 3, 5\}$.
- The set of elements that belongs to at least one of the sets $C_1, \ldots, C_k$ is $C_1 \cup C_2 \ldots \cup C_k$, also written $\cup_{i=1}^{k} C_i$.

Example

- $C \cup \phi =$
- $C \cup C =$
- If $C_1 \subset C_2$, then $C_1 \cup C_2 =$
- If $C_i = \{x : x \in [i-1, i]\}$, $i = 1, \dots, k$ then $\cup_{i=1}^k C_i = \{x : x \in [0, k]\}$.

Example (1.2.7)

$$C_k = \left\{ x : \frac{1}{k+1} \le x \le 1 \right\}, \qquad k = 1, 2, \dots.$$

Then

$$C_1 \cup C_2 \cup C_3 \cup \cdots = (0, 1].$$

Intersection

- The set of all elements that belong to each of the sets C_1 and C_2 is called the **intersection** of C_1 and C_2, written as $C_1 \cap C_2$.
- If $C_1 = \{1, 2, 3\}$, $C_2 = \{2, 3, 5\}$, then $C_1 \cup C_2 = \{2, 3\}$.
- If $C_1 = [0, 1]$ and $C_2 = [-1, 0]$ then $C_1 \cap C_2 = \{0\}$.
- If $C_i = (0, 1/i)$, $i = 1, \ldots, k$ then $\cap_{i=1}^{k} C_i = (0, 1/k)$.

Example (1.2.11)

Let

$$C_k = \left\{ x : 0 < x < \frac{1}{k} \right\}, \qquad k = 1, 2, \ldots .$$

Then

$$C_1 \cap C_2 \cap C_3 \cap \cdots = \emptyset.$$

Example

Use Venn diagrams to depict the sets $C_1 \cup C_2$, $C_1 \cap C_2$, $(C_1 \cup C_2) \cap C_3$ and $(C_1 \cap C_2) \cup C_3$.

Space & Complement

- The set of all elements under consideration is called the **space**, written as $\mathcal{C}$.
- Number of heads in tossing a coin ten times. The space is $\mathcal{C} = \{0, 1, \ldots, 10\}$.
- Let $\mathcal{C}$ be the sample space and C be its subset. The set that consists of all elements of $\mathcal{C}$ that are not elements of C is called the **complement** of C, written as C^c.
- Number of heads in tossing a coin ten times. If $C = \{0, 1, 2, 3, 4\}$ then $C^c = \{5, 6, 7, 8, 9, 10\}$.
  1. $C \cup C^c = \mathcal{C}$.
  2. $C \cap C^c = \emptyset$.
  3. $(C^c)^c = C$.

Important basic rules

1. **DeMorgan's laws**: Let $\mathcal{C}$ denote the space and suppose $C_1, C_2 \subset \mathcal{C}$. Then

$$A: \quad (C_1 \cap C_2)^c = C_1^c \cup C_2^c.$$
$$B: \quad (C_1 \cup C_2)^c = C_1^c \cap C_2^c.$$

2. **Distributive laws**: Let $\mathcal{C}$ denote the space and suppose $C_1, C_2, C_3 \subset \mathcal{C}$. Then

$$A: \quad C_1 \cup (C_2 \cap C_3) = (C_1 \cup C_2) \cap (C_1 \cup C_3).$$
$$B: \quad C_1 \cap (C_2 \cup C_3) = (C_1 \cap C_2) \cup (C_1 \cap C_3).$$

Set functions

- Usual function maps each point to a real number.
- Set function maps each set to a real number:

  More specifically, let $\mathcal{C}$ be a space and C be its subset. A mapping Q that assigns a value to the subset C (rather than an element x) is called a **set function**.

Example (1.2.18)

Let C be a set in one-dimensional space and let $Q(C)$ be equal to the number of points in C which correspond to positive integers. Then $Q(C)$ is a function of the set C. Thus:

- if $C = \{x : 0 < x < 5\}$, then $Q(C) = 4$.
- if $C = \{-2, -1\}$, then $Q(C) = 0$.
- if $C = \{x : -\infty < x < 6\}$, then $Q(C) = 5$.

Example (1.2.23)

Let C be a set in one-dimensional space and let

$$Q(C) = \int_C e^{-x} dx.$$

If $C = \{x : 0 \le x < \infty\}$, then

$$Q(C) = \int_0^\infty e^{-x} dx = 1.$$

If $C = \{x : 1 < x \le 3\}$, then

$$Q(C) = \int_1^3 e^{-x} dx = e^{-1} - e^{-3}.$$

Example (1.2.24)

Let C be a set in n-dimensional space and let

$$Q(C) = \int \underset{C}{\cdots} \int dx_1 \cdots dx_n.$$

If $C = \{(x_1, x_2, \ldots, x_n) : 0 \le x_1, x_2, \cdots, x_n \le 1\}$, then

$$Q(C) = \int_0^1 \int_0^1 \cdots \int_0^1 dx_1 dx_2 \cdots dx_n = 1.$$

If $C = \{(x_1, x_2, \ldots, x_n) : 0 \le x_1 \le x_2 \le \cdots \le x_n \le 1\}$, then

$$Q(C) = \int_0^1 \int_0^{x_n} \cdots \int_0^{x_3} \int_0^{x_2} dx_1 dx_2 \cdots dx_{n-1} dx_n = \frac{1}{n!}.$$

Chapter 1 Probability and Distributions

1.3 The Probability Set Function

σ-field

Definition (σ-field)

A collection of sets that is closed under complementation and countable union of its members is a *σ-field*. This collection of events is usually denoted by $\mathcal{B}$.

- The collection is also closed under countable intersections according to DeMorgan's Laws.

Probability set function

Definition (Probability set function)

Let $\mathcal{C}$ be a sample space and $\mathcal{B}$ be a σ-field defined on $\mathcal{C}$. Let P be a real-valued function defined on $\mathcal{B}$. Then P is a **probability function** if

1. $P(C) \geq 0$ for all $C \in \mathcal{B}$;
2. $P(\mathcal{C}) = 1$;
3. If $C_i \in \mathcal{B}$ $(i = 1, 2, \ldots)$ and $C_i \cap C_j = \Phi \ \forall i \neq j$, then $P(\cup_{i=1}^{\infty} C_i) = \sum_{i=1}^{\infty} P(C_i)$.

- A collection of events whose members are pairwise disjoint is said to be mutually exclusive.
- The collection is further said to be exhaustive if the union of its events is the sample space.
- These three axioms imply many properties of the probability set function. Let us see a few examples.

Basic results on probability functions

Theorem (1.3.1)

$P(C) = 1 - P(C^c)$.

Theorem (1.3.2)

$P(\phi) = 0$.

proof: Let $C = \emptyset$ and follows from $P(\mathcal{C}) = 1$.

Theorem (1.3.3)

If $C_1 \subset C_2$, *then* $P(C_1) \le P(C_2)$.

proof: Let $T = C_1 \cup (C_1^C \cap C_2)$. Notice that $C_2 = C_1 \cup T$ and $C_1 \cap T = \emptyset$.

Theorem (1.3.4)

$0 \leq P(C) \leq 1.$

proof: $0 = P(\emptyset) \leq P(C) \leq P(\mathcal{C})$.

Theorem (1.3.5)

For two arbitrary events C_1 and C_2, it holds that

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2).$$

Inclusion Exclusion Formula

For three arbitrary events C_1, C_2 and C_3, it holds that

$$\begin{aligned} P(C_1 \cup C_2 \cup C_3) = & P(C_1) + P(C_2) + P(C_3) \\ & - P(C_1 \cap C_2) - P(C_1 \cap C_3) - P(C_2 \cap C_3) \\ & + P(C_1 \cap C_2 \cap C_3). \end{aligned}$$

- Boole's inequality:

$$P(C_1) + P(C_2) + \ldots + P(C_k) \geq P(C_1 \cup C_2 \cup \ldots \cup C_k).$$

- Bonferroni's inequality:

$$P(C_1 \cap C_2) \geq P(C_1) + P(C_2) - 1.$$

If an experiment can result in any one of N different outcomes, and if exactly n of those outcomes correspond to event C, then the probability of event C is

$$P(C) = \frac{n}{N}$$

Example 1.3.2

An unbiased coin is to be tossed twice and the outcomes are in an ordered pair. Then,
$\mathcal{C} = \{(TT), (TH), (HT), (HH)\}$.

Write $C_1 = \{$the first toss results in a head$\}$,
$C_2 = \{$the second toss results in a tail$\}$.
Then, $P(C_1 \cup C_2) =?$

Method 1: $C_1 = \{HH, HT\}$, $C_2 = \{HT, TT\}$, so
$C_1 \cup C_2 = \{HH, HT, TT\}$. We see $P(C_1 \cup C_2) = 3/4$.

Method 2: $C_1 = \{HH, HT\}$, so $P(C_1) = 1/2$, $C_2 = \{HT, TT\}$, so $P(C_2) = 1/2$. We have $C_1 \cap C_2 = \{HT\}$, so
$P(C_1 \cap C_2) = 1/4$. Thus by Inclusion Exclusion Formula,
$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2) = 3/4$.

Counting rules

- Suppose we have two experiments, The first experiment results in m outcomes while the second results in n outcomes. The composite experiment (the first experiment followed by the second) has mn outcomes. This is called **the multiplication rule**.

- Let A be a set with n elements. Suppose we are interested in $k-$tuples whose components are elements of A. Then, by the extended multiplication rule, there are n^k such $k-$tuples.

Permutation

Suppose $k \leq n$ and we are interested in $k-$tuples whose components are **distinct** elements of A. Hence, by the multiplication rule, there are $\underline{n(n-1)\dots(n-(k-1))}$ such $k-$tuples with distinct elements.

Example

For the integers $1, 2, 3, 4, 5$, and ordered subsets of size 2 we have:

(1,2), **(1,3)**, **(1,4)**, **(1,5)**, $(2, 1)$, **(2,3)**, **(2,4)**, **(2,5)**, $(3, 1)$, $(3, 2)$, **(3,4)**, **(3,5)**, $(4, 1)$, $(4, 2)$, $(4, 3)$, **(4, 5)**, $(5, 1)$, $(5, 2)$, $(5, 3)$, $(5, 4)$.

The number of permutations of n things taken k at a time is

$$P_k^n = n(n-1) \cdot \ldots \cdot (n-j+1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

- Apply to previous example: $5!/2! = 20$.

Combination

- Suppose we have n objects, $a_1, \dots, a_n$. How many subsets of size k without regard for order can we choose from these objects?
  For the previous example we have 10.
- In general the number of *combinations* of n things taken k at a time is
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Example

Poker hand example: 52 cards, 5 cards in a hand.

1. Probability of any specific hand: $1/\binom{52}{5}$.
2. Probability that all cards are hearts: $\binom{13}{5}/\binom{52}{5}$.
3. Probability of a flush (all cards same suit): $\binom{4}{1}\binom{13}{5}/\binom{52}{5}$.
4. Probability of a full house (three kings and two queens): $\binom{4}{3}\binom{4}{2}/\binom{52}{5}$.

Theorem

The number of distinct permutations of n objects of which n_1 are of one kind, n_2 of a second kind, ..., n_k of a kth kind is

$$\frac{n!}{n_1!n_2!\ldots n_k!}.$$

Example

How many distinct permutations can be made form the word INTERNET?

- The letters are 8 letters in the word "INTERNET":

$$\begin{array}{ccccc} I & N & T & E & R \\ 1 & 2 & 2 & 2 & 1 \end{array}$$

- Number of different permutations:

$$\frac{8!}{1!2!2!2!1!} = 5040.$$

A different point of view

- If you put n distinct objects into k cells. And you want to put n_1 into cell 1, n_2 into cell 2, ..., n_k into cell k, where $n = n_1 + n_2 + \ldots + n_k$, then the answer is the same as

$$\frac{n!}{n_1!n_2!\ldots n_k!}$$

- Multiplication rule:

$$\begin{aligned}\binom{n}{n_1, n_2, \ldots, n_k} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\ldots\binom{n-n_1-\ldots n_{k-1}}{n_k} \\ &= \frac{n!}{n_1!n_2!\ldots n_k!}\end{aligned}$$

Chapter 1 Probability and Distributions

1.4 Conditional Probability and Independence

Motivation of conditional probability

- An experiment involves three tosses of a coin.
  Sample space consists of 8 possible outcomes, all equally likely:

$$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT.$$

- Define two events:
  $C_1 = \{$first toss results in a head$\}$,
  $C_2 = \{$at least two heads$\}$.
  What is the probability of event C_2?

$$P(C_2) =$$

- Suppose we know event C_1 occurs. Now what is the probability of event C_2?

$$P(C_2|C_1) =$$

(the conditional probability of C_2 given that C_1 occurs).

Definition of conditional probability

Definition. For two events C_1 and C_2, with C_1 satisfying $P(C_1) > 0$, the conditional probability of C_2 given C_1 is

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)}.$$

Properties of conditional probabilities:

1. $P(C_2|C_1) \geq 0$;
2. $P(C_2 \cup C_3 \cup \cdots |C_1) = P(C_2|C_1) + P(C_3|C_1) + \cdots$, provided that C_2, C_3, $\cdots$ are mutually disjoint;
3. $P(C_1|C_1) = 1$.

Example (1.4.2)

A bowl contains eight chips. Three of the chips are red and the remaining five are blue. Two chips are to be drawn successively, at random and without replacement. Compute the probability that the first draw results in a red chip (C_1) and the second draw results in a blue chip (C_2).

Solution:

$$P(C_1 \cap C_2) = P(C_1)P(C_2|C_1) = \frac{3}{8}\frac{5}{7} = \frac{15}{56}.$$
$$P(C_1 \cap C_2) = P(C_2)P(C_1|C_2) = \texttt{hard} \cdot \texttt{hard}.$$

The multiplication rule can be extended to three or more events.
For example, if $P(C_1 \cap C_2) > 0$, hence $P(C_1) > 0$, we have

$$P(C_1 \cap C_2 \cap C_3) = P(C_1)P(C_2|C_1)P(C_3|C_1 \cap C_2).$$

Example (1.4.4)

Four cards are drawn successively, at random and without replacement, from an ordinary deck of playing cards. Compute the probability of receiving a spade, a heart, a diamond, and a club, in that order?

Solution:

$$\frac{13}{52}\frac{13}{51}\frac{13}{50}\frac{13}{49}.$$

What is the probability of receiving a spade, a heart, a diamond, and a club in any order?

Solution:

$$4! \cdot \frac{13}{52}\frac{13}{51}\frac{13}{50}\frac{13}{49}.$$

Law of total probability

Suppose $C_1, \ldots, C_k$ form a partition of the sample space $\mathcal{C}$, i.e.

1. $C_1, \ldots, C_k$ are mutually exclusive
2. $C_1, \ldots, C_k$ are exhaustive, i.e. $P(C_1 \cup \ldots \cup C_k) = 1$

Then for any event $C \in \mathcal{B}$

$$P(C) = P(C|C_1)P(C_1) + P(C|C_2)P(C_2) + \cdots + P(C|C_k)P(C_k)$$

Proof:

$$\begin{aligned}
P(C) &= P[\mathcal{C} \cap C)] \\
&= P[(C_1 \cup C_2 \cup \ldots \cup C_k) \cap C)] \\
&= P[(C_1 \cap C) \cup \ldots \cup (C_k \cap C))] \\
&= P(C_1 \cap C) + \cdots + P(C_k \cap C) \\
&= P(C|C_1)P(C_1) + \cdots + P(C|C_k)P(C_k).
\end{aligned}$$

Bayes's theorem

Suppose $C_1, \ldots, C_k$ form a partition of the sample space $\mathcal{C}$. Then for any event $C \in \mathcal{B}$,

$$P(C_j|C) = \frac{P(C \cap C_j)}{P(C)} = \frac{P(C|C_j)P(C_j)}{\sum_{i=1}^{k} P(C|C_i)P(C_i)}$$

Example

On Friday evening 10% of the drivers in Iowa City are drunk. The probability that a drunk driver will be involved in a traffic accident is 0.01% and the probability that a sober driver will be involved in a traffic accident is 0.002%. If you read in the morning paper that a particular individual was involved in a traffic accident, what is the probability that this individual was drunk?

Solution:

Let $C =$ Accident occurs; $C_1 =$ Drunk; $C_2 =$ Sober. Then

$$\begin{aligned} P(C_1|C) &= \frac{P(C_1)P(C|C_1)}{P(C_1)P(C|C_1) + P(C_2)P(C|C_2)} \\ &= \frac{.10 \cdot .0001}{.10 \cdot .0001 + .9 \cdot .00002} = 0.357 \end{aligned}$$

Four inspectors at a film factory are stamping the expiration date on each package of a film at the end of the assembly line.

- John, 20% of the packages, fails to stamp 1 in 200,
- Tom, 60% of the packages, fails to stamp 1 in 100,
- Jeff, 15% of the packages, fails to stamp 1 in 90,
- Pat, 5% of the packages, fails to stamp 1 in 200.

If a customer complains that her package of film does not show the expiration date, what is the probability that it was inspected by John?

Solution:

- Let B_1 be the event that John inspected the package. Similarly, denote the event that the package was inspected by Tom, Jeff and Pat using B_2, B_3, B_4 respectively. Then B_1, B_2, B_3, B_4 form a partition of the sample space. Let F be the event that the package failed to be stamped.
- $P(B_1) = 0.2$, $P(B_2) = 0.6$, $P(B_3) = 0.15$, $P(B_4) = 0.05$. $P(F|B_1) = \frac{1}{200}$, $P(F|B_2) = \frac{1}{100}$, $P(F|B_3) = \frac{1}{90}$, $P(F|B_4) = \frac{1}{200}$.
- The probability that the package was inspected by John given the failure is

$$\begin{aligned} P(B_1|F) &= \frac{P(F|B_1)P(B_1)}{P(F|B_1)P(B_1) + P(F|B_2)P(B_2) + P(F|B_3)P(B_3) + P(F|B_4)P(B_4)} \\ &= \frac{(0.2)(1/200)}{(0.2)(1/200) + (0.6)(1/100) + (0.15)(1/90) + (0.05)(1/200)} \\ &= 0.112 \end{aligned}$$

Suppose 5 out of 10000 employees at Lawrence Livermore National Laboratory are spies. According to experts, a polygraph test has **sensitivity** of 0.91 and **specificity** of 0.94. Mathematically

$$\text{Prevalence} = P(S) = 0.0005$$
$$\text{Sensitivity} = P(+|S) = 0.91$$
$$\text{Specificity} = P(-|NS) = 0.94$$

1. Find the probability of a false positive test.
   Solution:
   $$P(+|NS) = 1 - P(-|NS) = 1 - 0.94 = 0.06$$
   The complement rule can be used for conditional probabilities *if* what is being conditioned on is held fixed (i.e. NS was held fixed in the above computation). In other words, $P(+|NS) \neq 1 - P(+|S)$.
2. Find the probability of a false negative test.
   Solution:
   $$P(-|S) = 1 - P(+|S) = 1 - 0.91 = 0.09$$

1. Suppose an employee is randomly selected. Find the probability that he/she tests positive for being a spy.
   Solution: By the law of total probability,

$$\begin{aligned} P(+) &= P(+|S)P(S) + P(+|NS)P(NS) \\ &= (0.91)(0.0005) + (0.06)(0.9995) = 0.060425. \end{aligned}$$

2. Given that a randomly selected employee tested positive, find the probability that he/she actually is a spy.
   Solution: By Bayes's theorem,

$$\begin{aligned} P(S|+) &= \frac{P(+ \cap S)}{P(+)} \\ &= \frac{P(+|S)P(S)}{P(+|S)P(S) + P(+|NS)P(NS)} \\ &= \frac{(0.91)(0.0005)}{0.060425} = 0.00753. \end{aligned}$$

Therefore, if an employee tests positive for being a spy, there is a 0.753% chance that he/she actually is a spy!

Definition

Two events C_1 and C_2 with $P(C_1) > 0$ and $P(C_2) > 0$ are **independent** if and only if $P(C_1 \cap C_2) = P(C_1)P(C_2)$.

Remark:

- Two events C_1, C_2 are independent if and only if $P(C_1 \cap C_2) = P(C_1)P(C_2)$, which is equivalent to $P(C_1|C_2) = P(C_1)$, when $P(C_2) > 0$, and also equivalent to $P(C_2|C_1) = P(C_2)$, when $P(C_1) > 0$.
- Two events are independent if the occurrence (or non-occurrence) of one event does not influence the likelihood of occurrence of the other event.

- If the two events C_1 and C_2 are independent, then the following pairs of events are also independent:
(i) C_1 and C_2^c; (ii) C_1^c and C_2; (iii) C_1^c and C_2^c.

Proof.

It suffices to show that

$$P(C_1^c \cap C_2) = P(C_1^c)P(C_2) = (1 - P(C_1))P(C_2),$$

where

$$\begin{aligned} P(C_2) &= P(\mathcal{C} \cap C_2) \\ &= P((C_1 \cup C_1^c) \cap C_2) \\ &= P[(C_1 \cap C_2) \cup (C_1^c \cap C_2)] \text{ (distributive law)} \\ &= P(C_1 \cap C_2) + P(C_1^c \cap C_2) \text{ (axiom of probability)} \\ &= P(C_1) \cdot P(C_2) + P(C_1^c \cap C_2). \text{ (independence)} \end{aligned}$$

$\square$

Mutual independence

Suppose now we have three events, C_1, C_2, and C_3. We say that they are **mutually independent** if

$$\begin{aligned} P(C_1 \cap C_2) &= P(C_1)P(C_2), \\ P(C_1 \cap C_3) &= P(C_1)P(C_3), \\ P(C_2 \cap C_3) &= P(C_2)P(C_3). \end{aligned}$$

and

$$P(C_1 \cap C_2 \cap C_3) = P(C_1)P(C_2)P(C_3).$$

More generally, we say the n events, C_1, C_2, ..., C_n, are mutually independent if for any collection of distinct integers, d_1, d_2, ..., d_k, from $\{1, 2, \ldots, n\}$, it holds that

$$P(C_{d_1} \cap C_{d_2} \cap \cdots \cap C_{d_k}) = P(C_{d_1})P(C_{d_2}) \cdots P(C_{d_k}).$$

Example

Suppose a circuit board contains 3 modules. The probability that the first module works properly is 0.98, while the second and third modules work properly with probability 0.95 and 0.92, respectively. Modules are independent.

1. Find the probability that all 3 modules work properly.

$$\begin{aligned} P(W_1 \text{ and } W_2 \text{ and } W_3) &\overset{\text{indep}}{=} P(W_1)P(W_2)P(W_3) \\ &= (0.98)(0.95)(0.92) = 0.8565 \end{aligned}$$

2. Find the probability that one or more modules work.

$$\begin{aligned} P(\text{one or more work}) &= 1 - P(0 \text{ work}) \\ &\overset{\text{indep}}{=} 1 - P(W_1^c)P(W_2^c)P(W_2^c) \\ &= 1 - (1-0.98)(1-0.95)(1-0.92) \end{aligned}$$

Chapter 1 Probability and Distributions

1.5 Random Variables

Definition of random variable

Definition A random variable X is a map from the sample space $\mathcal{C}$ to the real set $\mathcal{R}$. It assigns to each element $c \in \mathcal{C}$ one and only one value $X(c) = x$.

1. X induces a new sample space:

$$\mathcal{D} = \{x = X(c) : c \in \mathcal{C}\} \subset \mathbb{R}.$$

   Example 1: Coin flip: $\mathcal{C} = (T, H)$, and we create $X(T) = 0$, $X(H) = 1$.
   Example 2: $\mathcal{C} = \{c_1, \ldots, c_n\}$, students at the UIowa. $X(c_i) = c_i$'s height and $Y(c_i) = c_i$'s weight.

2 X induces a probability set function on $\mathcal{R}$:

$$P_X(B) = P\{c \in \mathcal{C} : X(c) \in B\} \quad \text{for } B \subset \mathbb{R}.$$

Consider $D \in \mathcal{D}$. We have
$P(X \in D) = P(\{c \in \mathcal{C} : X(c) \in D\})$.

Example

Flip a coin twice. Let $X =$ total number of heads. The sample space is $\mathcal{C} = \{HH, HT, TH, TT\}$ and all outcomes are equally likely, so

$$\begin{array}{cccc} x: & 0 & 1 & 2 \\ P(X = x): & \frac{1}{4} & \frac{2}{4} & \frac{1}{4} \end{array}$$

Discrete Random Variables

Discrete random variables

Definition

A random variable X is **discrete** if it can assume a finite or countably infinite (i.e. you can create a one-to-one correspondence with the positive integers) number of values, then X is a **discrete random variable**.

Example

1. Let X be the number of broken eggs in a dozen randomly selected eggs.
   Possible values for X are $x = 0, 1, \ldots, 12$, so X is discrete.
2. Let X be the number of accidents at an intersection in a year.
   Possible values for X are $x = 0, 1, 2, \ldots$, so X is discrete.
3. Flipping a coin until you obtain Head.
   Possible values for X are $x = 1, 2, \ldots$, so X is discrete.

Probability mass function

For the discrete random variables, the **probability mass function** (**pmf**)

$$p_X(d_i) = P(X = d_i) \quad i = 1, 2 \ldots,$$

determines the probability set function P_X.

The pmf $p_X(x)$ has the following properties,

1. $0 \le p_X(x) \le 1$,
2. $\sum_{x \in \mathcal{D}} p_X(x) = 1$,
3. $P[X \in A] = \sum_{x \in A} p_X(x)$.

Definition

The **support** of a discrete random variable is $\{x : P(X = x) > 0\}$, sometimes denoted S

Suppose we have a bowl with 1 chip labeled “1”, 2 chips labeled “2”, and 3 chips labeled “3”. Draw 2 chips *without* replacement. Let X = sum of the two draws.

1. Find the probability mass function (p.m.f.) of X.

$x:$	3	4	5	6
$p_X(x) = P(X = x):$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{6}{15}$	$\frac{3}{15}$

2. Find the probability that the sum of the two draws is 5 or more.

$$P(X \geq 5) = P(X = 5) + P(X = 6) = \frac{9}{15}.$$

3. Find the probability that the sum equals 6 given that the sum is 4 or more.

$$P(X = 6 | X \geq 4) = \frac{P(X = 6 \cap X \geq 4)}{P(X \geq 4)} = \frac{P(X = 6)}{P(X \geq 4)} = \frac{3}{13}.$$

4. Find the probability that the sum is 5 or less given that the sum is more than 4.

$$P(X \leq 5 | X > 4) = \frac{P(X \leq 5 \cap X > 4)}{P(X > 4)} = \frac{P(X = 5)}{P(X > 4)} = \frac{2}{3}.$$

Cumulative distribution function

Let X be a random variable. Then its **cumulative distribution function** (**cdf**) is defined by

$$F_X(x) = P_X((-\infty, x]) = P(X \leq x).$$

For a discrete random variable, this is a step function.

$$F_X(x_0) = \sum_{x \leq x_0} P_X(X = x).$$ (cdf is sum of pmf)

For the previous example

$$F(x) = \begin{cases} 0 & x < 3 \\ 2/15 & 3 \leq x < 4 \\ 6/15 & 4 \leq x < 5 \\ 12/15 & 5 \leq x < 6 \\ 1 & x \geq 6, \end{cases}$$

given the pmf

$$\begin{array}{rcccc} x: & 3 & 4 & 5 & 6 \\ p_X(x) = P(X = x): & \frac{2}{15} & \frac{4}{15} & \frac{6}{15} & \frac{3}{15} \end{array}$$

Continuous Random Variables

Continuous random variables

Definition

A random variable is a **continuous random variable** if its cumulative function $F_X(x)$ is a continuous function for all $x \in \mathbb{R}$.

The **cumulative distribution function (cdf)** defined for the discrete random variables can be used to describe continuous random variables too.

Example

Let X denote a real number chosen at random between 0 and 1. Since any number can be chosen equally likely, it is reasonable to assume

$$P_X[(a,b)] = b - a, \qquad \text{for } 0 < a < b < 1. \tag{1}$$

Describe its cdf.

Probability density function

- The notion of **probability mass function (pmf)** defined for discrete random variables does NOT work here: $F_X(x) - F_X(x-) = 0 \; \forall x \in \mathbb{R}$, so $P(X = x) = 0 \; \forall x \in \mathbb{R}$.
- Instead, for the continuous case, if there is a non-negative function f_X such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad \text{for all } -\infty < x < \infty, \quad \text{(cdf is integral of pdf)}$$

- We call f_X the **probability density function (pdf)** of the random variable X. It is easy to see that the following relation usually holds:

$$f_X(x) = F_X'(x), \qquad \text{for all } -\infty < x < \infty.$$

A **pdf** $f_X(x)$ always has the properties

1. $f(x) \geq 0 \ \forall x \in \mathbb{R}$.
2. $\int_{\mathbb{R}} f(x)dx = 1$.
3. If $A \subset \mathbb{R}$, then $P(A) = \int_A f(x)dx$. In other word,

$$P(a < X < b) = \int_a^b f(x)dx.$$

Definition

The **support** of a continuous random variable is $\{x : f_X(x) > 0\}$, sometimes denoted S

Let

$$f(x) = \begin{cases} k(1-x^2), & -1 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Find

1. the constant k such that $f(x)$ is a pdf.
2. $P(-0.5 < X < 0.5)$.
3. $P(X \leq 0.1)$.
4. $P(X > 0.7 | X > 0.1)$.
5. Evaluate $F(x)$.

Solution:

1. We must have
$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{1} k(1-x^2)\,dx = k\int_{-1}^{1} 1 - x^2\,dx$$
Thus $k \cdot \frac{4}{3} = 1$ and $k = \frac{3}{4}$.
2. By the definition of PDF, we have
$$P(-0.5 < X < 0.5) = \int_{-0.5}^{0.5} \frac{3}{4}(1-x^2)\,dx = \frac{11}{16} = 0.6875.$$
3. $$P(X \le 0.1) = P(X < 0.1) = \int_{-1}^{0.1} \frac{3}{4}(1-x^2)\,dx = 0.57475.$$
4. For r.v.'s, the convention is to use comma instead of $\cap$.
$$P(X > 0.7 | X > 0.1) = \frac{P(X > 0.7, X > 0.1)}{P(X > 0.1)}$$
$$= \frac{P(X > 0.7)}{P(X > 0.1)} = \frac{\int_{0.7}^{1} \frac{3}{4}(1-x^2)\,dx}{\int_{0.1}^{1} \frac{3}{4}(1-x^2)\,dx} = 0.143.$$

Percentile & quartile

Definition

Let X be a continuous-type random variable with pdf $f(x)$ and cdf $F(x)$. The $(100p)$**th percentile** is a number π_p such that the area under $f(x)$ to the left of π_p is p. That is,

$$p = \int_{-\infty}^{\pi_p} f(x)dx = F(\pi_p).$$

The 50th percentile is called the **median** ($m = \pi_{0.50}$). The 25th and 75th percentiles are called the **first and third quartiles**, respectively, denoted by $q_1 = \pi_{0.25}$ and $q_3 = \pi_{0.75}$. Of course, the median $m = \pi_{0.50} = q_2$ is also called the second quartile.

Example

Let Y be a continuous random variable with probability density function given by

$$f(y) = \begin{cases} 3y^2 & 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find the first, second and third quartiles.

Solution:

$$\int_0^{\pi_{0.25}} 3y^2 dy = 1/4, \quad \pi_{0.25} = 1/\sqrt[3]{4}.$$

$$\int_0^{\pi_{0.5}} 3y^2 dy = 1/2, \quad \pi_{0.5} = 1/\sqrt[3]{2}.$$

$$\int_0^{\pi_{0.75}} 3y^2 dy = 3/4, \quad \pi_{0.75} = \sqrt[3]{3}/\sqrt[3]{4}.$$

Basic Properties of CDF (for both discrete and continuous r.v.)

Theorem 1.5.2

Let X be a random variable with cdf F_X. Then,

$$P_X(a, b] = P(a < X \leq b) = F_X(b) - F_X(a).$$

Proof.

It holds that

$$\{-\infty < X \leq b\} = \{-\infty < X \leq a\} \cup \{a < X \leq b\}.$$

The theorem follows from the third axiom of the probability since the two sets on the right-hand side are disjoint. $\square$

Example 1.5.7 & 1.5.8

Determine the constant c and the probability $P(2 < X \leq 4)$ in the following questions:

1. X has a pmf

$$p_X(x) = \begin{cases} cx & x = 1, 2, \ldots, 10 \\ 0 & \text{otherwise.} \end{cases}$$

2. X has a pdf

$$f_X(x) = \begin{cases} cx & 0 < x < 10 \\ 0 & \text{otherwise.} \end{cases}$$

Example

Consider an urn which contains balls each with one of the numbers $1, 2, 3, 4$ on it. Suppose there are i balls with the number i for $i = 1, 2, 3, 4$. Suppose one ball is drawn at random. Let X be the number on the ball.

(a) Determine the pmf of X;

(b) Compute $P(X \leq 3)$;

(c) Determine the cdf of X.

Theorem 1.3.6 Continuity of Probability

For an increasing sequence of events $\{C_n\}$, define its limit as $\lim_{n\to\infty} C_n = \bigcup_{n=1}^{\infty} C_n$. It holds that

$$\lim_{n\to\infty} P(C_n) = P(\lim_{n\to\infty} C_n) = P\left(\bigcup_{n=1}^{\infty} C_n\right).$$

Symmetrically, for an decreasing sequence of events $\{C_n\}$, define its limit as $\lim_{n\to\infty} C_n = \bigcap_{n=1}^{\infty} C_n$. It holds that

$$\lim_{n\to\infty} P(C_n) = P(\lim_{n\to\infty} C_n) = P\left(\bigcap_{n=1}^{\infty} C_n\right).$$

Theorem 1.5.1

Let X be a random variable with cdf F.

(a) If $a < b$, then $F(a) \leq F(b)$.
(b) $\lim_{x \to -\infty} F(x) = 0$.
(c) $\lim_{x \to \infty} F(x) = 1$.
(d) $\lim_{x \searrow x_0} F(x) = F(x_0)$.

Proof.

(a) If $a < b$, then $\{X \leq a\} \subset \{X \leq b\}$. The result then follows from the monotonicity of the probability.

(b) If $\{x_n\}$ is an decreasing sequence such that $x_n \to -\infty$, then $C_n = \{X \leq x_n\}$ is decreasing with $\emptyset = \cup_{n=1}^{\infty} C_n$. From the continuity of probability theorem,

$$\lim_{n \to -\infty} F(x_n) = P\left(\cap_{n=1}^{\infty} C_n\right) = P(\emptyset) = 0.$$

□

Theorem 1.5.1 (cont'd)

Let X be a random variable with cdf F.

(c) $\lim_{x\to\infty} F(x) = 1$.

(d) $\lim_{x\searrow x_0} F(x) = F(x_0)$ (right continuous).

Proof.

(c) If $\{x_n\}$ is an increasing sequence such that $x_n \to \infty$, then $C_n = \{X \leq x_n\}$ is increasing with $\{X \leq \infty\} = \cup_{n=1}^{\infty} C_n$. From the continuity of probability theorem,

$$\lim_{n\to\infty} F(x_n) = P\left(\cup_{n=1}^{\infty} C_n\right) = 1.$$

(d) Let $\{x_n\}$ be any decreasing sequence of real numbers such that $x_n \downarrow x_0$. Then the sequence of sets $\{C_n\}$ is decreasing and $\cap_{n=1}^{\infty} C_n = \{X \leq x_0\}$. The continuity of probability theorem implies that

$$\lim_{n\to\infty} F(x_n) = P\left(\cap_{n=1}^{\infty} C_n\right) = F(x_0).$$

Theorem 1.5.3

For any random variable,

$$P(X = x) = F_X(x) - F_X(x-),\ \forall x \in \mathbb{R}.$$

Proof.

For any x, we have $\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right]$. By the continuity of the probability function,

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

Summary

- Random variable X is a function from a sample space $\mathcal{C}$ into the real numbers $\mathbb{R}$.
- Every random variable is associated with a cdf:

$$F_X(x) = P_X(X \leq x) \text{ for } -\infty < x < \infty.$$

- $F_X(x)$ is defined for all x, not just those in its domain $\mathcal{D}$.
- Notation wise, random variables will always be denoted with uppercase letters and the realized values of the variable will be denoted by the corresponding lowercase letters. Thus the random variable X can take the value x.
- We say a random variable X is **discrete** if $F_X(x)$ is a step function: $F(x_0) = \sum_{x \leq x_0} P(X = x)$.
- We say a random variable X is **continuous** if $F_X(x)$ is a continuous function: $F(x_0) = \int_{-\infty}^{x_0} f(x)dx$.

Chapter 1 Probability and Distributions

1.6 Discrete Random Variables

Review of discrete random variable

$\mathcal{D} = \{x = X(c) : c \in \mathcal{C}\}$ is either finite or countable.
("$\mathcal{D}$ is countable" means that there is a one-to-one correspondence between $\mathcal{D}$ and the positive integers.)

The **support** of a discrete random variable X is $\{x : p_X(x) > 0\} =$
(In English: the support of X consists of all points x such that $p_X(x) > 0$.)

Example: geometric distribution

Consider a sequence of independent flips of a coin, each resulting in a head with probability p and a tail with probability $q = 1 - p$. Let the random variable X be the number of tails before the first head appears. Determine the pmf of X and the probability that X is even.

Solution:

$$P(X = x) = q^x p, \quad x = 0, 1, 2, \ldots$$

$$\begin{aligned}\sum_{k=0}^{\infty} P(X = 2k) &= p \sum_{k=0}^{\infty} (q^2)^k \\ &= p/(1 - q^2).\end{aligned}$$

Example 1.6.2: hypergeometric distribution

A lot consists of m good fuses and n defective fuses. Choose k, $k \le \min\{m, n\}$, fuses at random from the lot. Let the random variable X be the number of defective fuses among the k. Determine the pmf of X.

Solution:

$$p_X(x) = \begin{cases} \dfrac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}} & \text{for } x = 0, 1, \ldots, k, \\ 0 & \text{otherwise.} \end{cases}$$

Transformation of discrete random variables

Let X be a discrete random variable with pmf $p_X(x)$ and support $\mathcal{D}_{\mathcal{X}} = \{x : p_X(x) > 0\}$. Suppose we are interested in $Y = g(X)$. Then Y is also a random variable, and we want to determine its pmf.

Let $\mathcal{D}_Y = \{y : y = g(x)$ for some $x \in \mathcal{D}_X\}$. We have

$$\begin{aligned} p_Y(y) &= P(Y = y) \\ &= P(g(X) = y) \\ &= \sum_{x : g(x) = y} p_X(x) \\ &= p_X(g^{-1}(y)) \quad \text{if } g \text{ is one to one} \\ &= P(X \in g^{-1}(y)), \end{aligned}$$

where $g^{-1}(y) = \{x : g(x) = y\}$.

Example Consider the geometric random variable X

pmf: $P(X = x) = q^x p, \; x = 0, 1, 2, \ldots$.

1. Let Y be the number of flips needed to obtain the first head. Then $Y = X + 1$.
   $Y = g(X) = X + 1$ and $X = g^{-1}(Y) = Y - 1$.

   Determine the pmf of Y.
   Solution:

$$p_Y(y) = p_X(g^{-1}(y)) = q^{y-1}p.$$

2. Let $Y = (X - 2)^2$. Determine the pmf of Y.
   Solution:

$$p_Y(y) = \begin{cases} p_X(2) & \text{if } y = 0, \\ p_X(1) + p_X(3) & \text{if } y = 1, \\ p_X(0) + p_X(4) & \text{if } y = 4, \\ p_X(\sqrt{y} + 2) & \text{if } y \geq 9, 16, 25 \ldots \end{cases}$$

Chapter 1 Probability and Distributions

1.7 Continuous Random Variables

Review of continuous random variable

- Recall that a random variable is **continuous** if its cdf $F_X(x)$ is a continuous function for all $x \in \mathbb{R}$. Hence, if the random variable X is continuous, then $P(X = x) = F_X(x) - F_X(x-) = 0, \quad$ for all $x \in \mathbb{R}$. This means that there is no point of discrete mass.
- Recall that a nonnegative function f_X is a pdf of the random variable X if

$$\int_{-\infty}^{x} f_X(t)dt = F_X(x), \qquad \text{for all } -\infty < x < \infty,$$

- The **support** of a continuous random variable X is $\{x : f_X(x) > 0\} = \mathcal{D}$.
  (In English: the support of X consists of all points x such that $f_X(x) > 0$.)

Example 1.7.1

Suppose we select a point at random in the interior of a circle of radius 1. Let X be the distance of the selected point from the origin. Determine the cdf and pdf of X and the probability that the selected point falls in the ring with radius $1/4$ and $1/2$.

Solution:

$$P(X \leq x) = x^2,$$

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \leq x < 1, \\ 1, & x \geq 1, \end{cases}$$

$$f_X(x) = \begin{cases} 2x, & 0 \leq x < 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$P(1/4 < X \leq 1/2) = \int_{1/4}^{1/2} 2t dt = 3/16.$$

Example 1.7.2 (Exponential Distribution)

Let the random variable X be the time in seconds between incoming telephone calls at a busy switchboard. Suppose that X has a pdf

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

This is an exponential distribution with parameter/rate $\lambda > 0$. Knowing that the parameter $\lambda = 1/4$, compute the probability that the time between successive phone calls exceeds 4 seconds.
Solution:

$$P(X > 4) = \int_4^\infty \frac{1}{4} e^{-x/4} dx = e^{-1} = 0.3679.$$

Transformation of Continuous R.V.

Sometimes we know the distribution of a continuous random variable X. We are interested in the distribution of a random variable Y which is a **transformation** (function) of X, say $Y = g(X)$, and we want to determine the distribution of Y.

Example 1.7.3. Let X be the random variable with the pdf

$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Determine the cdf and pdf of $Y = X^2$.
Solution:

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = y.$$

$$f_Y(y) = \begin{cases} 1 & 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.7.1: Transformation of continuous r.v.

Assume that X is a continuous random variable with pdf $f_X(x)$ and support S_X. Suppose $Y = g(X)$, where g is one-to-one. Then,

$$f_Y(y) = f_X(h(y))|h'(y)|, \quad \text{where } h = g^{-1}.$$

Proof: Suppose that g (and thus h) is monotone increasing. Then

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[X \leq h(y)] = F_X[h(y)],$$

Therefore,

$$f_Y(y) = F'_Y(y) = F'_X[h(y)] \cdot h'(y) = f_X[h(Y)]h'(y) = f_X[h(Y)]\frac{dx}{dy}.$$

Note that since h is increasing $h'(y)$ is positive on S_Y.
$(h'(y) = |h'(y)|)$.

Proof (cont'd)

Suppose that g (and therefore h) is monotone decreasing. Then,

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[X \geq h(y)] = 1 - F_X[h(y)]$$

Therefore,

$$f_Y(y) = F'_Y(y) = -F'_X[h(y)] \cdot h'(y) = f_X[h(Y)](-h'(y))$$

Note that since h is decreasing $h'(y)$ is negative on S_Y. ($-h'(y) = |h'(y)|$).

We refer to $dx/dy = J$ as the **Jacobian** of the transformation.

Example

Assume that $f(x) = 1$ when $0 < x < 1$ and 0 otherwise. Therefore $S_X = (0, 1)$.

1. Let $Y = X^2$. $g(x) = x^2$ is a one-to-one function on $(0, 1)$. $h(y) = y^{1/2}$, and $h'(y) = \frac{1}{2}y^{-1/2}$. Therefore,

$$f_Y(y) = f_X(h(y))|h'(y)| = 1 \cdot \frac{1}{2}y^{-1/2} = \frac{1}{2}y^{-1/2},$$

when $0 < y < 1$. ($S_Y = (0, 1)$).

2. Suppose now that $Y = g(X) = -\log(X)$. Then $h(y) = \exp(-y)$, and $|h'(y)| = \exp(-y)$. Therefore

$$f_Y(y) = f_X(h(y))|h'(y)| = 1 \cdot \exp(-y) = \exp(-y) \quad y > 0.$$

You can verify that $\int_0^\infty f_Y(y)dy = 1$.

Extended theorem: Transformation of continuous r.v.

Assume X is a continuous random variable with pdf $f_X(x)$ and support S_X. Suppose $Y = g(X)$ and the support has a partition $A_1, \ldots A_k$. Further, suppose there exist functions $g_1(x), \ldots, g_k(x)$ such that

- $g(x) = g_i(x)$, for $x \in A_i$,
- $g_i(x)$ is monotone on each $x \in A_i$,
- define $\mathcal{Y}_i = \{y = g_i(x), \forall x \in A_i\}$, then $\mathcal{Y}_1 = \mathcal{Y}_2 = \ldots = \mathcal{Y}_k$.

Then

$$F_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right|, & y \in \mathcal{Y}, \\ 0 & \text{otherwise.} \end{cases}$$

Example

- Assume that $f(x) = 1/2$ when $-1 < x < 1$ and 0 otherwise. Therefore $S_X = (-1, 1)$.
- Let $Y = X^2$ and $g(x) = x^2$. Then $g_1(x) = x^2$ is a one-to-one function on $(-1, 0)$ and $g_2(x) = x^2$ is one-to-one on $(0, 1)$. We see $\mathcal{Y}_i = \{y : y = g_i(x), x \in A_i\} = (0, 1)$.
- Then $h_1(y) = g_1^{-1}(y) = -y^{1/2}$, and $h_1'(y) = -\frac{1}{2}y^{-1/2}$. $h_2(y) = g_2^{-1}(y) = y^{1/2}$, and $h_2'(y) = \frac{1}{2}y^{-1/2}$.
- Therefore,

$$\begin{aligned} f_Y(y) &= f_X(h_1(y))|h_1'(y)| + f_X(h_2(y))|h_2'(y)| \\ &= \frac{1}{2} \cdot \frac{1}{2}y^{-1/2} + \frac{1}{2} \cdot \frac{1}{2}y^{-1/2} \\ &= \frac{1}{2}y^{-1/2}, \end{aligned}$$

when $0 < y < 1$. ($S_Y = (0, 1)$).

Mixture of Discrete and Continuous R.V.

Example 1.7.6

Study the cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x+1}{2} & 0 \le x < 1 \\ 1 & 1 \le x. \end{cases}$$

- What is $P(-3 < X \le 1/2)$.
- What is $P(X = 0)$.
- What is $P(-3 < X \le 0)$.
- What is $P(-3 < X < 0)$.

It always holds from the definition of cdf that

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a).$$

Example 1.7.7

Let X equals the size of a wind loss in millions of dollars, and suppose it has the cdf

$$F_X(x) = \begin{cases} 0 & -\infty < x < 0 \\ 1 - \left(\frac{10}{10+x}\right)^3 & 0 \le x < \infty. \end{cases}$$

If losses beyond $\$10,000,000$ are reported as 10, then the cdf of this censored distribution is

$$F_Y(y) = \begin{cases} 0 & -\infty < y < 0 \\ 1 - \left(\frac{10}{10+y}\right)^3 & 0 \le y < 10 \\ & \\ 1 & 10 \le y < \infty. \end{cases}$$

where $Y = \min(X, 10)$,

- This is an example of a mixed continuous and discrete random variable. The particular example chosen is known as **censoring** because it creates the discrete part by lumping one end of the distribution into a single point.
- The continuous part of the random variable has the same pdf as the pdf for X on $(0, 10)$, i.e.

$$\frac{3 \cdot 10^3}{(y+4)^4} 1_{(0,10)}(y).$$

- The discrete part has the pmf

$$P(Y = 10) = 1 - F_X(10) = \left(\frac{10}{20}\right)^3 = \frac{1}{8}.$$

Chapter 1 Probability and Distributions

1.8 Expectation of a Random Variable

Definition of expectation

Let X be a random variable.

- If X is continuous with pdf $f(x)$ then the **expectation** of X is

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

  provided that $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$.
  We say the expectation does not exist if $\int_{-\infty}^{\infty} |x| f(x) dx = \infty$.

- If X is discrete with pmf $p(x)$ then the **expectation** of X is

$$\mathrm{E}(X) = \sum_x x p(x).$$

  provided that $\sum_x |x| p(x) < \infty$.
  We say the expectation does not exist if $\sum_x |x| p(x) = \infty$.

- Sometimes the expectation of X is called the mathematical expectation of X, the expected value of X, or the mean of X. We often denote $E(X)$ by μ ($\mu = E(X)$).

Example

Suppose we have a bowl with 1 chip labeled "1", 2 chips labeled "2", and 3 chips labeled "3". Draw 2 chips *without* replacement. Let $X =$ sum of the two draws. The pmf (derived before) is as follows:

$$\begin{array}{rcccc} x: & 3 & 4 & 5 & 6 \\ p_X(x) = P(X = x): & \frac{2}{15} & \frac{4}{15} & \frac{6}{15} & \frac{3}{15} \end{array}$$

We now find the expected value (mean) of the random variable X.

$$\mu = \sum_{x=3}^{6} xp(x) = 3\left(\frac{2}{15}\right) + 4\left(\frac{4}{15}\right) + 5\left(\frac{6}{15}\right) + 6\left(\frac{3}{15}\right) = 4.667.$$

If the random variable X was observed many times and the realizations were recorded, the expected value, μ, describes **the mean of the observed realizations**. If I repeatedly drew 2 chips from the bowl without replacement and recorded the sum of the two draws, the mean of the recorded sums would equal $\mu = 4.667$.

Exponential distribution

Let X be a random variable that follows the exponential distribution with parameter λ, that is,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Compute its expectation.
Solution:

$$\begin{aligned} \mathrm{E}X &= \int_0^\infty \lambda x e^{-\lambda x} dx \\ &= -\int_0^\infty x d e^{-\lambda x} \\ &= -x e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= \frac{1}{\lambda}. \end{aligned}$$

Two formulations of exponential distribution

1. The parameter λ is defined as **rate**:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

   In this case, $EX = 1/\lambda$.

2. The parameter λ is defined as **scale**:

$$f_X(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{1}{\lambda} x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

   In this case, $EX = \lambda$.

3. The first case is used in this course.

Expectation of a Function of X

Let X be a random variable and let $Y = g(X)$ for some function g.
We can calculate $\mathrm{E}Y = \mathrm{E}g(X)$ in two different ways.
Suppose X is discrete and $\sum_{x \in \mathcal{D}_X} \left|g(x)\right| p_X(x) < \infty$, then

1. $$\mathrm{E}g(X) = \sum_{x \in \mathcal{D}_X} g(x)\, p_X(x).$$
2. Let $p_Y(y)$ be the pmf of Y. Then
$\mathrm{E}g(X) = \mathrm{E}Y = \sum_{y \in \mathcal{D}_Y} y\, p_Y(y)$.

Suppose X is continuous and $\int_{-\infty}^{\infty} \left|g(t)\right| f_X(t)dt < \infty$, then

1. $$\mathrm{E}g(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$
2. Let $Y = g(X)$ and let $f_Y(y)$ be the pdf of Y. Then
$\mathrm{E}g(X) = \mathrm{E}Y = \int_{-\infty}^{\infty} y f_Y(y) dy$.

Assume X has a pdf $f(x) = 1$, for $0 < x < 1$. Let $Y = -\log X$.
What is $\mathrm{E}Y$?
Solution:

$$\mathrm{E}Y = \int_0^1 -\log x dx = -[x \log x - x]|_0^1 = 1.$$

For a certain ore samples the proportion X of impurities per sample is a random variable with density function given by

$$f(x) = \begin{cases} \frac{3}{2}x^2 + x & 0 \le x \le 1, \\ & \\ 0 & \text{elsewhere.} \end{cases}$$

The dollar value of each sample is $Y = 5 - 0.5X$. Find the mean of X and Y.

Solution:

$$\mathrm{E}X = \int_0^1 x\left(\frac{3}{2}x^2 + x\right)dx = \frac{17}{24} = 0.708.$$

$$\mathrm{E}Y = \int_0^1 (5 - 0.5x)\left(\frac{3}{2}x^2 + x\right)dx = 5 - \frac{17}{48} = 4.646.$$

Linearity of expectation

Theorem (1.8.2)

Let $g_1(X)$ *and* $g_2(X)$ *be functions of a random variable* X*.*
Suppose the expectations of $g_1(X)$ *and* $g_2(X)$ *exist.*
Then for any constants k_0*,* k_1 *and* k_2*, the expectation of* $k_0 + k_1 g_1(X) + k_2 g_2(X)$ *exists and it is given by*

$$E\left[k_0 + k_1 g_1(X) + k_2 g_2(X)\right] = k_0 + k_1 E\left[g_1(X)\right] + k_2 E\left[g_2(X)\right].$$

Corollary: If $g(X) = a + bX$, i.e. g is a linear function, then

$$E[g(X)] = a + bE(X) = g[E(X)]$$

Example

Assume X has a pdf $f(x) = 1$, for $0 < x < 1$.

$$E(X + 3X^2) = E(X) + 3E(X^2) = \frac{1}{2} + 3 \cdot \frac{1}{3} = \frac{3}{2}$$

Warning: If g is a nonlinear function, then in general $E[g(X)] \neq g[E(X)]$. For the previous example

$$E(X^2) = \frac{1}{3} \neq (E(X))^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

Which one is larger? Jensen inequality...

Example

Let X have a pdf $f(x) = 3x^2$, $0 < x < 1$, zero elsewhere. Consider a random rectangle whose sides are X and $(1 - X)$. Determine the expected value of the area of the rectangle.

Solution:

$$\int_0^1 3x^2 \cdot x(1-x)dx = 0.15.$$

Suppose X is a random variable such that $\mathrm{E}\big(X^2\big) < \infty$. Consider the function

$$h(b) = \mathrm{E}\left[\big(X - b\big)^2\right].$$

The value of b that minimizes $h(b)$ is $\mathrm{E}X$.

Let X be a continuous random variable with cdf $F(x)$. Determine the expectation of $F(X)$.

Chapter 1 Probability and Distributions

1.9 Some Special Expectations

Mean and variance

Definition (Mean)

Let X be a random variable whose expectation exists. The **mean value** μ of X is defined to be $\mu = \mathrm{E}(X)$.

Definition (Variance)

Let X be a random variable with finite mean μ. Then the **variance** of X is defined to be $\sigma^2 = \mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$.
The positive square root $\sigma = \sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X.

Computation of $\mathrm{Var}(X)$:

$$\mathrm{Var}(X) = \mathrm{E}X^2 - (\mathrm{E}X)^2,$$

which is because E is a linear operator.

Coin-flipping example

Define $X = 1$ if it is head and $X = 0$ if it is tail.
Assume $P(X = 1) = p$ and then $P(X = 0) = 1 - p$.
Find the mean and variance of X.

Solution:

$$\begin{aligned} \mathrm{E}(X) &= 1 \cdot p + 0 \cdot (1-p) = p, \\ \mathrm{E}(X^2) &= 1^2 \cdot p + 0^2 \cdot (1-p) = p, \\ \sigma^2 &= p - p^2 = p(1-p). \quad \sigma = \sqrt{p(1-p)}. \end{aligned}$$

Assume $a > 0$ and let

$$f_X(x) = \begin{cases} 1/a & 0 < x < a \\ 0 & \text{otherwise.} \end{cases}$$

Solution:

$$\begin{aligned} \mu &= \int_0^a x \frac{1}{a} dx = \frac{a}{2}, \\ \mathrm{E}(X^2) &= \int_0^a x^2 \frac{1}{a} dx = \frac{a^2}{3}, \\ \sigma^2 &= \frac{a^2}{3} - \frac{a^2}{4} = \frac{a^2}{12}. \end{aligned}$$

Example 1.9.2

If X has the pdf

$$f_X(x) = \begin{cases} \frac{1}{x^2} & 1 < x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

then the mean value of X does not exist:

$$\begin{aligned} \int_1^\infty x \cdot \frac{1}{x^2} dx &= \int_1^\infty \frac{1}{x} dx \\ &= \infty. \end{aligned}$$

Linear transformation

- Suppose that $Y = a + bX$. Then,
  1. The mean of Y is $\mathrm{E}(Y) = \mu_Y = a + b\mathrm{E}(X) = a + b\mu_X$.
  2. The variance of Y is

$$\mathrm{Var}(Y) = \mathrm{E}(Y - \mu_Y)^2 = \mathrm{E}(b^2(X - \mu_X)^2) = b^2\sigma_X^2.$$

     In fact,

$$\mathrm{Var}(Y) = \mathrm{Var}(a + bX) = \mathrm{Var}(bX) = b^2\mathrm{Var}(X).$$

  3. The standard deviation of Y is $|b|\sigma_X$.
- Let the random variable X have a mean μ and a standard deviation σ. Show that

$$\mathrm{E}\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] = 1.$$

Expectation of non-negative random variables

- Let X be a random variable of the discrete type with pmf $p(x)$, $x = 0, 1, 2, \ldots$. It holds that

$$\mathrm{E}(X) = \sum_{x=0}^{\infty} \left[1 - F(x)\right].$$

- Let X be a continuous random variable with pdf $f(x)$. Suppose that $f(x) = 0$ for $x < 0$. It holds that

$$\mathrm{E}(X) = \int_0^{\infty} \left[1 - F(x)\right] dx.$$

Proof for discrete case

$$\begin{aligned}
\sum_{x=0}^{\infty}[1 - F(x)] &= \sum_{x=0}^{\infty}[1 - P(X \le x)] = \sum_{x=0}^{\infty} P(X > x) \\
&= \sum_{x=0}^{\infty} \sum_{t=x+1}^{\infty} P(X = t) \\
&= \sum_{x=0}^{\infty} \sum_{t=0}^{\infty} \mathbf{1}_{t>x} P(X = t) \\
&= \sum_{t=0}^{\infty} \sum_{x=0}^{\infty} \mathbf{1}_{x<t} P(X = t) \quad \Big(\text{interchangable due to absolute convergence}\Big) \\
&= \sum_{t=0}^{\infty} t P(X = t) = \mathrm{E}(X).
\end{aligned}$$

Proof for continuous case

In this 4000-level class, we only prove

$$\mathrm{E}(X) = \int_0^b [1 - F(x)]\, dx,$$

where the support of X is $(0, b)$ and $b < \infty$, although the equality holds for $b = \infty$.

Proof.

$$\begin{aligned}
\int_0^b [1 - F(x)]dx &= (x - xF(x))\,\Big|_0^b + \int_0^b xf(x)dx \\
&= \int_0^b xf(x)dx \\
&= \mathrm{E}(X).
\end{aligned}$$

$\square$

- The m'**th raw moment** of X is defined as $\mathrm{E}(X^m)$ if the expectation exists.
- The m'**th central moment** of X is defined as $\mathrm{E}(X-\mu)^m$ if the expectation exists.

Example

1. Coin flipping: $P(X=1)=p$,

$$\mathrm{E}(X^m) = 1^m \cdot p + 0^m(1-p) = p.$$

2. Assume $a > 0$ and let

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

$$\mathrm{E}(X^m) = \int_0^a x^m \frac{1}{a} dx = \frac{a^m}{m+1}.$$

Moment generating function

- Let X be a random variable. Assume there is a positive number h such that $E[\exp(tX)]$ exists for all $t \in (-h, h)$.
- The **moment generating function (mgf)** of X is defined to be the function

$$M_X(t) = \mathrm{E}\left(e^{tX}\right), \qquad -h < t < h.$$

Example

1. Assume $P(X = 0) = 1 - p$ and $P(X = 1) = p$.

$$M_X(t) = \mathrm{E}[e^{tX}] = p \cdot e^t + (1-p) \cdot e^0 = pe^t + 1 - p, \text{ for } -\infty < t < \infty.$$

2.

$$f_X(x) = \begin{cases} 1/a & 0 < x < a \\ 0 & \text{otherwise.} \end{cases}$$

$$M_X(t) = \mathrm{E}[e^{tX}] = \frac{1}{a}\int_0^a e^{tx}dx = \frac{1}{at}e^{tx}|_0^a = \frac{e^{at} - 1}{at},$$

when $t \neq 0$ and $M_X(t) = 1$ when $t = 0$.

Properties of mgf

1. It is always true that $M_X(0) = 1$.
2. The moments of X can be found (or "generated") from the successive derivatives of $M_X(t)$.

$$M'_X(0) = \mathrm{E}(X), \quad M''_X(0) = E(X^2), \quad M_X^{(n)} = E(X^n).$$

- We have $M'_X(0) = \mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \mu$, since

$$M'(t) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)dx = \int_{-\infty}^{\infty} \frac{d}{dt} e^{tx} f(x)dx = \int_{-\infty}^{\infty} xe^{tx} f(x)dx$$

- We then see

$$M''(0) = \mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \mu^2 + \sigma^2,$$

so $\sigma^2 = \mathrm{E}(X^2) - \mu^2 = M''(0) - [M'(0)]^2$.

The pdf of X is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Find the mgf of X and use it to find the mean and the variance.
Solution:

The mgf is given as follows,

$$M(t) = \mathrm{E}e^{tx} = \int_0^\infty e^{tx} f(x)dx = \frac{\lambda}{\lambda - t},\ t < 1,$$

and

$$M'(t) = \frac{\lambda}{(\lambda - t)^2},$$
$$M''(t) = \frac{2\lambda}{(\lambda - t)^3}.$$

Thus

$$\mu = M'(0) = 1/\lambda,\ \sigma^2 = M''(0) - \mu^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2.$$

Theorem 1.9.1

Theorem

Let X and Y be random variables with moment generating functions M_X and M_Y, respectively, existing in open intervals about 0. If $M_X(t) = M_Y(t)$ for all t in an interval counting $t = 0$, then X and Y have identical probability distributions.

Example

Suppose X is a random variable of the continuous type with mgf

$$M(t) = \frac{1}{1-3t}, \qquad t < \frac{1}{3}.$$

Identify its distribution.

Solution:
The random variable X follows the exponential distribution with rate $\lambda = 1/3$.

Characteristic function

- Distributions may not have mgf.
- Can you show the mgf of Cauchy distribution does not exist?

$$f(x) = \frac{1}{\pi}\frac{1}{x^2+1}, \ -\infty < x < \infty$$

- **Characteristic function**: (not in exams)

$$\phi(t) = \mathrm{E}(e^{itX}), \quad \text{for an arbitrary real value } t.$$

- **Every** distribution has a **unique** characteristic function, and $\mathrm{E}(X) = i\phi'(0)$ and $\mathrm{E}(X^2) = -\phi''(0)$.
- Uniqueness of characteristic function is due to the uniqueness of Fourier transform. In fact, uniqueness of mgf comes from the uniqueness of Laplacian transform.

Chapter 1 Probability and Distributions

1.10 Important Inequalities

Chebyshev's inequality

Motivation

- The variance of a random variable tells use something about the variability of the observations about the mean.
- If a random variable has a small variance or standard deviation, we would expect most of the values to be grouped around the mean.
- For any random variable, the probability between any two values symmetric about the mean should be related to the standard deviation.

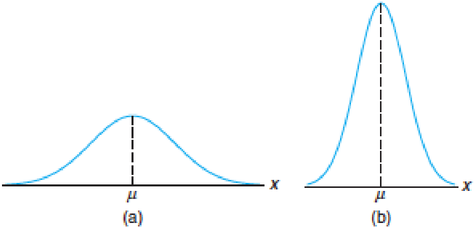

Figure 4.2: Variability of continuous observations about the mean.

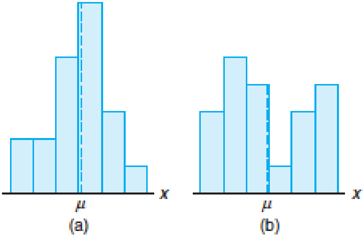

Figure 4.3: Variability of discrete observations about the mean.

Chebyshev's Theorem

Theorem (1.10.3)

Suppose for a random variable X, $\mathrm{E}(X^2)$ exists, then for any constant $k > 0$,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

Example

A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find the lower bounds of

1. $P(-4 < X < 20)$,
2. $P(|X - 8| \geq 6)$.

Solution:

$$P(-4 < X < 20) = P(8 - 4 \times 3 < X < 8 + 4 \times 3) \geq 15/16.$$

$$\begin{aligned} P(|X - 8| \geq 6) &= 1 - P(|X - 8| < 6) \\ &= 1 - P(-6 < X - 8 < 6) \\ &= 1 - P(8 - 2 \times 3 < X < 8 + 2 \times 3) \leq 1/4. \end{aligned}$$

Comments on Chebyshev's Theorem

- Chebyshev's theorem holds for any distribution of observations.
- For this reason, the results are usually weak. The value given by the theorem is a lower bound only.
- We know the probability of a random variable falling within two standard deviations will be no less than $3/4$, but we never know how much more it might actually be, unless we can determine exact probabilities.
- Chebyshev's theorem is thus called ***distribution-free*** result. **The results will be less conservative when specific distributions are known.**

Example

Compute $P(\mu - 2\sigma < X < \mu + 2\sigma)$, where X has the density function

$$f(x) = \begin{cases} 6x(1-x), & 0 < x < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and compare with the result given in Chebyshev's theorem.

Solution:

$\mu = E(X) = 6\int_0^1 x^2(1-x)\ dx = 0.5$, $E(X^2) = 6\int_0^1 x^3(1-x)\ dx = 0.3$, which imply $\sigma^2 = 0.3 - (0.5)^2 = 0.05$ and $\sigma = 0.2236$. Hence,

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P(0.5 - 0.4472 < X < 0.5 + 0.4472) \\ &= P(0.0528 < X < 0.9472) = 6\int_{0.0528}^{0.9472} x(1-x)\ dx = 0.9839, \end{aligned}$$

compared to a probability of at least 0.75 given by Chebyshev's theorem.

Jensen's inequality

Definition 1.10.1

A function ϕ defined on an interval (a, b) is said to be a **convex** function if for all x and y in (a, b) and all $0 < \gamma < 1$,

$$\phi\left[\gamma x + (1-\gamma)y\right] \leq \gamma\phi(x) + (1-\gamma)\phi(y).$$

We say ϕ is **strictly convex** if the above inequality is strict.

If ϕ is a (strictly) convex function, then $-\phi$ is a (strictly) **concave** function.

Theorem 1.10.4

If ϕ is twice differentiable on (a, b), then

(a) (first-order condition) ϕ is convex if and only if

$$\phi(x_1) \geq \phi(x_2) + \phi'(x_2)(x_1 - x_2).$$

(b) ϕ is strictly convex if and only if

$$\phi(x_1) > \phi(x_2) + \phi'(x_2)(x_1 - x_2).$$

(c) (second-order condition) ϕ is convex if and only if $\phi''(x) \geq 0$ for all $x \in (a, b)$;

(d) ϕ is strictly convex if and only if $\phi''(x) > 0$ for all $x \in (a, b)$.

The second-order condition is usually used to check the convexity, provided that the second-order derivative exists.

Jensen's inequality

Let ϕ be a convex function on an open interval I, let X be a random variable whose support is contained in I. If $\mu = \mathrm{E}(X)$ exists then

$$\phi\left[\mathrm{E}(X)\right] \leq \mathrm{E}\left[\phi(X)\right].$$

The inequality reverses if ϕ is a concave function.

Proof.

By the first-order condition,

$$\phi(x) \geq \phi(\mu) + \phi'(\mu)(x - \mu).$$

Then taking expectations of both sides leads to the result. The inequality is strict if ϕ is strictly convex. $\square$

Example

We have $\mu^2 < E(X^2)$ as $\phi(t) = t^2$ is strictly convex.

Example

Let X be a positive random variable. Argue that

1. $$\mathrm{E}\left(\frac{1}{X}\right) > \frac{1}{\mathrm{E}(X)}.$$

2. $$\mathrm{E}(\sqrt{X}) < \sqrt{\mathrm{E}\,(X)}.$$

Solution:

1. When $x > 0$, $\phi''(x) = 2x^{-3} > 0$, so $\phi(x) = 1/x$ is strictly convex.
   The result follows from Jensen's inequality.
2. When $x > 0$, $\phi''(x) = -1/(4x^{3/2}) < 0$, so $\phi(x) = \sqrt{x}$ is strictly concave.
   The result follows from Jensen's inequality.