Discrete Random Variables

1.1 Random Variables

In many areas of science we are interested in quantifying the **probability** that a certain outcome of an experiment occurs. We can use a **random variable** to identify numerical events that are of interest in an experiment. In this way, a random variable is a theoretical representation of the physical or experimental process we wish to study. More precisely, a random variable is a quantity without a fixed value, but which can assume different values depending on how likely these values are to be observed; these likelihoods are probabilities.

To quantify the probability that a particular value, or set of values (called an **event**), occurs, we use a number between 0 and 1. A probability of 0 implies that the event *cannot* occur, whereas a probability of 1 implies that the event *must* occur. Any value in the interval (0, 1) means that the event will only occur some of the time. Equivalently, if an event occurs with probability p, then this means there is a $p(100)\%$ chance of observing this event.

Conventionally, we denote random variables by capital letters, and particular values that they can assume by lowercase letters. So we can say that X is a random variable that can assume certain particular values x with certain probabilities.

We use the notation $\Pr(X = x)$ to denote the probability that the random variable X assumes the particular value x. The range of values x for which this expression makes sense is of course dependent on the possible values of the random variable X. We distinguish between two key cases.

If X can assume only finitely many or countably many values, then we say that X is a **discrete random variable**. Saying that X can assume only *finitely many or countably many* values means that we should be able to *list* the possible values for the random variable X. If this list is finite, we can say that X may take any value from the list $x_1, x_2, ..., x_n$, for some positive integer n. If the list is (countably) infinite, we can list the possible values for X as $x_1, x_2, ...$. This is then a list without end (for example, the list of all positive integers).

We summarize the basic notions of a discrete random variable:

1. A discrete random variable X is a quantity that can assume any value x from a discrete list of values with a certain probability.
2. The probability that the discrete random variable X assumes the particular value x is denoted by $\Pr(X = x)$. This collection of probabilities, along with all possible values x, is the **probability distribution** of the random variable X.
3. A discrete list of values is any collection of values that is finite or countably infinite (i.e. can be written in a list).

This terminology is in contrast to a **continuous random variable**, where the values the random variable can assume are given by a continuum of values. For example, we could define a random variable that can take any value in the interval [1,2]. The values X can assume are then any real number in [1,2]. We will discuss continuous random variables in detail in the second chapter. For now, we deal strictly with discrete random variables.

We state a few facts that should be intuitively obvious for probabilities in general. Namely, the chance of some particular event occurring should always be nonnegative and no greater than 100%. Also, the chance that *something* happens should be certain. From these facts, we can conclude that the chance of witnessing a particular event should be 100% less the chance of seeing *anything but* that particular event.

Discrete Probability Rules
1. Probabilities are numbers between 0 and 1 inclusive: $0 \leq \Pr(X = x_k) \leq 1$ for all k
2. The sum of all probabilities for a given experiment (random variable) is equal to one: $$\sum_k \Pr(X = x_k) = 1$$
3. The probability of an event is 1 minus the probability that any other event occurs: $$\Pr(X = x_n) = 1 - \sum_{k \neq n} \Pr(X = x_k)$$

Example: Tossing a Fair Coin Once

If we toss a coin into the air, there are only two possible outcomes: it will land as either "heads" (H) or "tails" (T). If the tossed coin is a "fair" coin, it is equally likely that the coin will land as tails or as heads. In other words, there is a 50% chance (1/2 probability) that the coin will land heads, and a 50% chance (1/2 probability) that the coin will land tails. Notice that the sum of these probabilities is 1 and that each probability is a number in the interval [0,1].

We can define the random variable X to represent this coin tossing experiment. That is, we define X to be the discrete random variable that takes the value 0 with probability 1/2 and takes the value 1 with probability 1/2. Notice that with this notation, the experimental event that "we toss a fair coin and observe heads" is the same as the theoretical event that "the random variable X is observed to take the value 0"; i.e. we identify the number 0 with the outcome of "heads", and identify the number 1 with the outcome of "tails". We say that X is a **Bernoulli random variable** with parameter 1/2 and can write $X \sim \text{Ber}(1/2)$.

Example: Tossing a Fair Coin Twice

Similarly, if we toss a fair coin two times, there are four possible outcomes. Each outcome is a sequence of heads (H) or tails (T):

- HH
- HT
- TH
- TT

Because the coin is fair, each outcome is equally likely to occur. There are 4 possible outcomes, so we assign each outcome a probability of 1/4.

Equivalently, we notice that for any of the four possible events to occur, we must observe two distinct events from two separate flips of a fair coin. So for example, to observe the sequence HH, we must flip a fair coin once and observe H, then flip a fair coin again and observe H once again. (We say that these two events are **independent** since the outcome of one event has no effect on the outcome of the other.) Since the probability of observing H after a flip of a fair coin is 1/2, we see that the probability of observing the sequence HH should be $(1/2) \times (1/2) = 1/4$.

Observe that again, all of our probabilities sum to 1, and each probability is a number on the interval [0, 1]. Just as before, we can identify each outcome of our experiment with a numerical value. Let us make the following assignments:

- HH -> 0
- HT -> 1
- TH -> 2
- TT -> 3

This assignment defines a numerical discrete random variable Y that represents our coin tossing experiment. We see that Y takes the value 0 with probability 1/4, 1 with probability 1/4, 2 with probability 1/4, and 3 with probability 1/4. Using our general notation to describe this probability distribution, we can summarize by writing

$\Pr(Y = k) = 1/4,$ for $k = 0, 1, 2, 3.$

Notice that with this notation, the experimental event that "we toss two fair coins and observe first tails, then heads" is the same as the theoretical event that "the random variable Y is observed to take the value 2". We say that Y is a **uniform discrete random variable** with parameter 4 since Y takes each of its four possible values with equal, or uniform, probability. To denote this distributional relationship, we can write $Y \sim$ Uniform(4).

1.2 Probability Basics

Random Variables and their Observed Values

We commonly use uppercase letters to denote random variables, and lowercase letters to denote particular values that our random variables can assume.

For example, consider a six-sided die, pictured below.
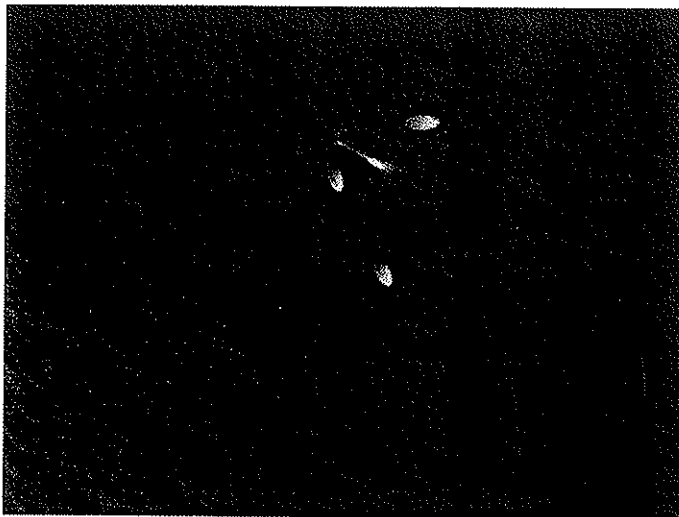
We could let X be the random value that gives the value observed on the upper face of the six-sided die after a single roll. Then if x denotes a particular value of the upper face, the expression $X = x$ becomes well-defined. Specifically, the notation $X = x$ signifies the event that the *random variable X* assumes the *particular value x*. For the six-sided die example, x can be any integer from 1 to 6. So the expression $X = 4$ would express the event that a random roll of the die would result in observing the value 4 on the upper face of the die.

Probabilities

We have already defined the notation $\Pr(X = x)$ to denote the probability that a random variable X is equal to a particular value x. Similarly, $\Pr(X \leq x)$ would denote the probability that the random variable X is less than or equal to the value x.

Notation
$\Pr(a \leq X \leq b)$ denotes the probability that the random variable X lies between values a and b, inclusively.

With this notation, it now makes sense to write, for example, $\Pr(X > a)$, the probability that a random variable assumes a particular value strictly greater than a. Similarly, we can make sense of the expressions $\Pr(X < b)$, $\Pr(X \neq x)$, $\Pr(X = x_1 \text{ or } X = x_2)$, among others.

Notice that this notation allows us to do a kind of algebra with probabilities. For example, we notice the equivalence of the following two expressions: $\Pr(X \geq a \text{ and } X < b) = \Pr(a \leq X < b)$. An important consequence of this symbolism is the following:

Probabilities of Complimentary Events
$\Pr(X = x) = 1 - \Pr(X \neq x)$
$\Pr(X > x) = 1 - \Pr(X \leq x)$
$\Pr(X \geq x) = 1 - \Pr(X < x)$

Notice that the first identity is simply a restatement of Discrete Probability Rule #3 from the previous section.

These three identities are simple consequences of our notation and of the fact that the sum of *all* probabilities must always equal 1 for any random variable. The events $X = x$ and $X \neq x$ are called **complimentary** because exactly one of the events must take place; i.e. both events cannot occur simultaneously, but one of the two must occur. The other expressions above also define complimentary events.

For discrete random variables, we also have the identity:

Algebra of Disjoint Events
If $a \neq b$, then $\Pr(X = a \text{ or } X = b) = \Pr(X = a) + \Pr(X = b)$

Six-Sided Die Example

Using our six-sided die example above, we have the random variable X which represents the value we observe on the upper face of the six-sided die after a single roll. Then the probability that X is equal to 5 can be written as:

$$\Pr(X = 5) = \frac{1}{6}$$

Using our identities for complimentary events and for disjoint events, we find that the probability that X is equal to 1, 2, 3 or 4 can be computed as:

$$\begin{aligned}
\Pr(1 \leq X \leq 4) &= \Pr(X = 1, \text{ or } X = 2, \text{ or } X = 3, \text{ or } X = 4) \\
&= 1 - \Pr(X = 5, \text{ or } X = 6) \\
&= 1 - [\Pr(X = 5) + \Pr(X = 6)] \\
&= 1 - \left(\frac{1}{6} + \frac{1}{6}\right) \\
&= \frac{2}{3}
\end{aligned}$$

Notice that $X \sim$ Uniform(6); i.e. X has a uniform distribution on the integers from 1 to 6. Indeed, the probability of observing any one of these integer values (the value on the upper face of the rolled die) is the same for any value. Thus, X must be a uniform random variable.

1.3 The Probability Mass Function

Usually we are interested in experiments where there is more than one outcome, each having a possibly different probability. The **probability mass function** of a discrete random variable is simply the collection of all these probabilities.

Probability Mass Function

The probability mass function (PMF) of a discrete random variable X provides the probabilities $\Pr(X = x)$ for all possible values of x. This function can be represented in a table, graph or formula.

Example: Different Colored Balls

Although it is usually necessary to define random variables that assume numerical values, this need not always be the case. Suppose that a box contains 10 balls:

- 5 of the balls are red
- 2 of the balls are green
- 2 of the balls are blue
- 1 ball is yellow

Suppose we take one ball out of the box. Let X be the random variable that represents the color of the ball. As 5 of the balls are red, and there are 10 balls in total, the probability that a red ball is drawn from the box is $\Pr(X = \text{Red}) = 5/10 = 1/2$. [Note: this random variable does not have numerical outcomes, but we could easily fix this by assigning different numbers to the different colors we could observe.]

Similarly, there are 2 green balls, so the probability that X is green is 2/10. Similar calculations for the other colors yield the probability mass function of X given by the following table.

Ball Color	Probability
red	5/10
green	2/10
blue	2/10
yellow	1/10

Example: A Six-Sided Die

Consider again the experiment of rolling a six-sided die. A six-sided die can land on any of its six faces, so that a single experiment has six possible outcomes.

For a "fair die", we anticipate getting each of the results with an equal probability, i.e. if we were to repeat the same experiment many times, we would expect that, on average, the six possible events would occur with similar frequencies (we say that such events are uniformly distributed).

There are six possible outcomes: 1, 2, 3, 4, 5, or 6. The probability mass function could be given by the following table.

Outcome	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

The PMF could also be given by the equation $\Pr(D = k) = 1/6$, for $k = 1, 2, 3, \ldots, 6$, where D denotes the random variable associated to rolling a fair die once. Thus we see that discrete uniform random variables have PMFs which are particularly easy to represent.

Example: Test Scores

Suppose that in a class of 10 people the grades on a test are given by 30, 30, 30, 60, 60, 80, 80, 80, 90, 100. Suppose a test is drawn from the pile at random and the score X is observed. We would like to calculate the probability mass function for the randomly drawn test score.

Looking at the test scores, we see that out of 10 grades:

- the grade 30 occurred 3 times
- the grade 60 occurred 2 times
- the grade 80 occurred 3 times
- the grade 90 occurred 1 time
- the grade 100 occurred 1 time

This tells us the probability mass function of the randomly chosen test score X which we present formally in the following table.

Grade, x_k	Probability, $\Pr(X = x_k)$
30	3/10
60	2/10
80	3/10
90	1/10
100	1/10

We have described the PMF in a table, but an equivalent representation could be given in a graph that plots the possible outcomes of X on the horizontal axis, and the probabilities associated to these outcomes on the vertical axis. Below is the graph of the probability mass function for the random variable X.

Probability Distribution for X
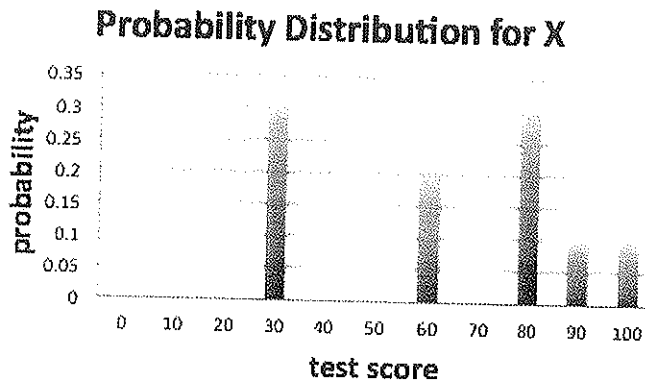
1.4 The Cumulative Distribution Function

Generally speaking, for any random variable X, we define the **cumulative distribution function** (CDF) of X as follows:

Cumulative Distribution Function
The cumulative distribution function (CDF) of a random variable X is denoted by $F(x)$, and is defined to be the function $$F(x) = \Pr(X \leq x).$$

In other words, the cumulative distribution function for a random variable at x gives the probability that the random variable X is less than or equal to that number x.

Given a *discrete* random variable and its associated probability mass function, the definition of the cumulative distribution function can be rewritten using our identity for the probability of disjoint events (see Section 1.02).

Cumulative Distribution Function of a Discrete Random Variable
If X is a discrete random variable, the cumulative distribution function (CDF) of X can be written as $$F(x) = \sum_{k=1}^{n} \Pr(X = x_k)$$ where x_n is the largest possible value of X that is less than or equal to x.

Note that in this formula for CDFs of discrete random variables, we always have $n \leq N$, where N is the number of possible outcomes of X.

Notice also that the CDF of a discrete random variable will remain constant on any interval of the form $[x_n, x_{n+1})$. That is, $F(x) = F(x_n) = \sum_{k=1}^{n} \Pr(X = x_k)$ for any $x \in [x_n, x_{n+1})$.

The following properties are immediate consequences of our definition of a random variable and the probability it associates to an event.

Properties of the CDF
• $0 \leq F(x) \leq 1$ for all x
• $\lim_{x \to -\infty} F(x) = 0$
• $\lim_{x \to +\infty} F(x) = 1$
• $F(x)$ is a *nondecreasing* function of x

Recall that a function $f(x)$ is said to be *nondecreasing* if $f(x_1) \leq f(x_2)$ whenever $x_1 < x_2$.

Example: Rolling a Single Die

If X is the random variable we associated previously with rolling a fair six-sided die, then we can easily write down the CDF of X.

We already computed that the PMF of X is given by $\Pr(X = k) = 1/6$ for $k = 1,2,...,6$. The CDF can be computed by summing these probabilities sequentially; we summarize as follows:

- $\Pr(X \leq 1) = 1/6$
- $\Pr(X \leq 2) = 2/6$
- $\Pr(X \leq 3) = 3/6$
- $\Pr(X \leq 4) = 4/6$
- $\Pr(X \leq 5) = 5/6$
- $\Pr(X \leq 6) = 6/6 = 1$

Notice that $\Pr(X \leq x) = 0$ for any $x < 1$ since X cannot take values less than 1. Also, notice that $\Pr(X \leq x) = 1$ for any $x > 6$. Finally, note that the probabilities $\Pr(X \leq x)$ are constant on any interval of the form $[k, k+1)$ as required.

Example: Rolling Two Dice

Suppose that we have two fair six-sided dice, one yellow and one red as in the image below.
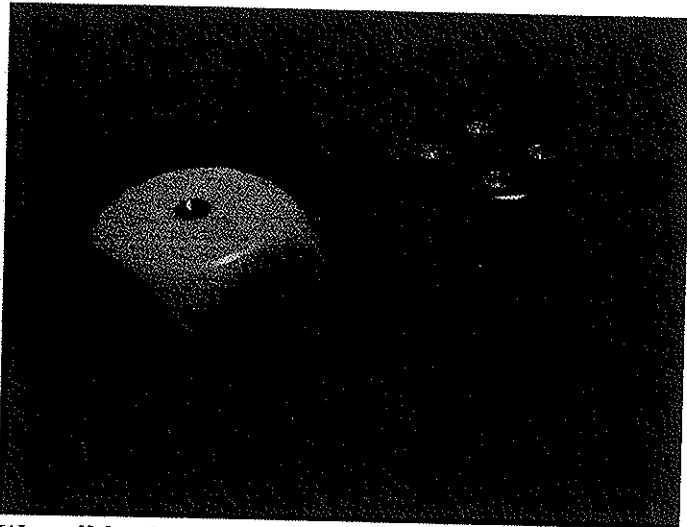
We roll both dice at the same time and add the two numbers that are shown on the upward faces.

Let Y be the discrete random variable associated to this sum.

1. How many possible outcomes are there? That is, how many different values can Y assume?

2. How is Y distributed? That is, what is the PMF of Y?

3. What is the probability that Y is less than or equal to 6?

4. What is the CDF of Y?

Solution

Part 1)

There are 6 possible values we can observe of each die. The two dice are rolled independently (i.e. the value on one of the dice does not affect the value on the other die), so we see that there are $6 \times 6 = 36$ different outcomes for a single roll of the two dice. Notice that all 36 outcomes are distinguishable since the two dice are different colors. So we can distinguish between a roll that produces a 4 on the yellow die and a 5 on the red die with a roll that produces a 5 on the yellow die and a 4 on the red die.

However, we are interested in determining the number of possible outcomes for the *sum* of the values on the two dice, i.e. the number of different values for the random variable Y. The smallest this sum can be is $1 + 1 = 2$, and the largest is $6 + 6 = 12$. Clearly, Y can also assume any value in between these two extremes; thus we conclude that the possible values for Y are 2,3,...,12.

Part 2)

To determine the probability distribution for Y, first consider the probability that the sum of the dice equals 2. There is only one way that this can happen: both dice must roll to 1. There are 36 distinguishable rolls of the dice, so the probability that the sum is equal to 2 is 1/36.

The other possible values of the random variable Y and their corresponding probabilities can be calculated in a similar fashion. Some of these are listed in the table below.

Outcome (Yellow, Red)	Sum = Yellow + Red	Probability
(1,1)	2	1/36
(1,2), (2,1)	3	2/36
(1,3), (2,2), (3,1)	4	3/36
(1,4), (2,3), (3,2), (4,1)	5	4/36
(1,5), (2,4), (3,3), (4,2), (5,1)	6	5/36
. . .	. . .	. . .
(6,6)	12	1/36

The probability mass function of Y is displayed in the following graph.

Probability Distribution for X
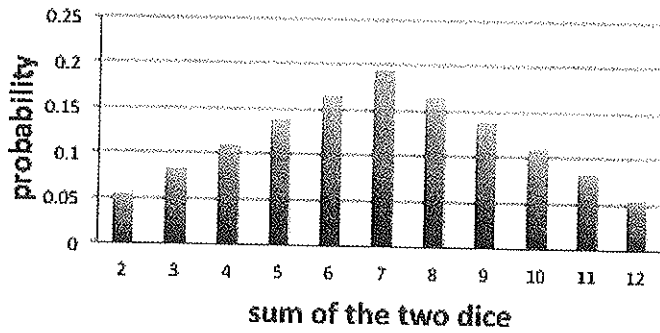
Alternatively, if we let $p_k = \Pr(Y = k)$, the probability that the random sum Y is equal to k, then the PMF can be given by a single formula:

$$p_k = \frac{6 - |k - 7|}{36} \text{ if } k = 2, 3, \ldots, 12$$

Part 3)

The probability that the sum is less than or equal to 6 can be written as $\Pr(Y \leq 6)$, which is equal to $F(6)$, the value of the cumulative distribution function $F(y)$ of Y at $y = 6$. Using our identity for probabilities of disjoint events, we calculate

$$\Pr(Y \leq 6) = F(6) = \sum_{k=1}^{6} p_k = 0 + \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{15}{36} = \frac{5}{12}$$

Part 4)

To find the CDF of Y in general, we need to give a table, graph or formula for $F(k) = \Pr(Y \leq k)$ for any given k. Using our table for the PMF of Y, we can easily construct the corresponding CDF table:

$Y = k$	$F(k) = \Pr(Y \leq k)$
2	1/36
3	3/36
4	6/36
5	10/36
6	15/36
. . .	. . .
12	36/36 = 1

This table defines a step-function starting at 0 for $y < 2$ and increasing in steps to 1 for $y \geq 12$. Notice that the CDF is constant over any half-closed integer interval from 2 to 12. For example, $F(y) = 3/36$ for all y in the interval [3,4).
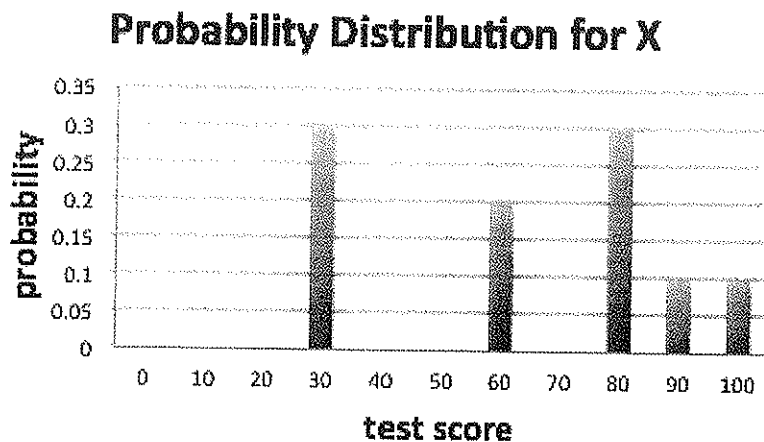
Example: Test Scores

Consider the example of selecting a test score from a given collection that we explored in the previous section: in a class of 10 people, grades on a test were 30, 30, 30, 60, 60, 80, 80, 80, 90, 100. Let X be the score of a randomly drawn test from this collection.

1. Calculate the probability that a test drawn at random has a score less than or equal to 80.
2. Calculate the probability that a test drawn at random has a score less than or equal to x_n, where $x_n = 0, 10, 20, 30, \ldots, 100$.

Solution

Part 1)

Recall the probability mass function, calculated earlier:

Probability Distribution for X

Let p_k be the probability that the score of a randomly drawn test is $x_k = 10k$. So, for example:

- p_0 is the probability that a randomly drawn test score is 0
- p_1 is the probability that a randomly drawn test score is 10
- p_2 is the probability that a randomly drawn test score is 20
- p_3 is the probability that a randomly drawn test score is 30

and so on. Values for each of these probabilities are given in the above bar graph. Notice that many of these probabilities are zero.

The probability that a test drawn at random has a score of no greater than 80 is exactly the value of the CDF of X at $x = 80$; i.e.,

$$\Pr(X \le 80) = F(80)$$

$$= \sum_{k=0}^{8} p_k$$

$$= p_0 + p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8$$

$$= 0 + 0 + 0 + \frac{3}{10} + 0 + 0 + \frac{2}{10} + 0 + \frac{3}{10}$$

$$= \frac{8}{10}$$

$$= \frac{4}{5}$$

The color blue was used in the above calculation to highlight nonzero probabilities.

Because of the sample space of our experiment, if the randomly selected grade is to be less than or equal to 80, then this grade can only be 30, 60, or 80. Intuitively, the probability that a randomly selected test has a grade of 30, 60, or 80 is the *sum* of the probabilities that the score is one of these possibilities, which we note is in agreement with our identity concerning probabilities of disjoint events from Section 1.02.

Part 2)

Now we want to calculate the probability that a test drawn at random has a score less than or equal to $x_k = 10k$ for $k = 0, 1, ..., 10$. Again, we identify this as simply finding the value of the CDF of X at each of these x_k values.

$$\Pr(X \leq 0) = F(0) = \sum_{k=0}^{0} p_k = p_0 = 0$$

Similarly, $F(0) = F(10) = F(20) = 0$. $F(30)$ is non-zero:

$$\Pr(X \leq 30) = F(30) = \sum_{k=0}^{3} p_k = 0 + 0 + 0 + \frac{3}{10}$$

Notice that $F(40)$ is equal to $F(30)$, since $p_4 = 0$.

Other values of F are calculated in the same way using the definition of the cumulative distribution function. The following table contains the values of the CDF of X for $x_k = 0, 10, 20, 30, ... 100$.

k	x_k	$F(x_k)$
0	0	0
1	10	0
2	20	0
3	30	0.3
4	40	0.3
5	50	0.3
6	60	0.5
7	70	0.5
8	80	0.8
9	90	0.9
10	100	1.0

Collectively, our calculations give the CDF of the random variable X. This cumulative distribution function is graphed in the figure below.

Cumulative Distribution for X
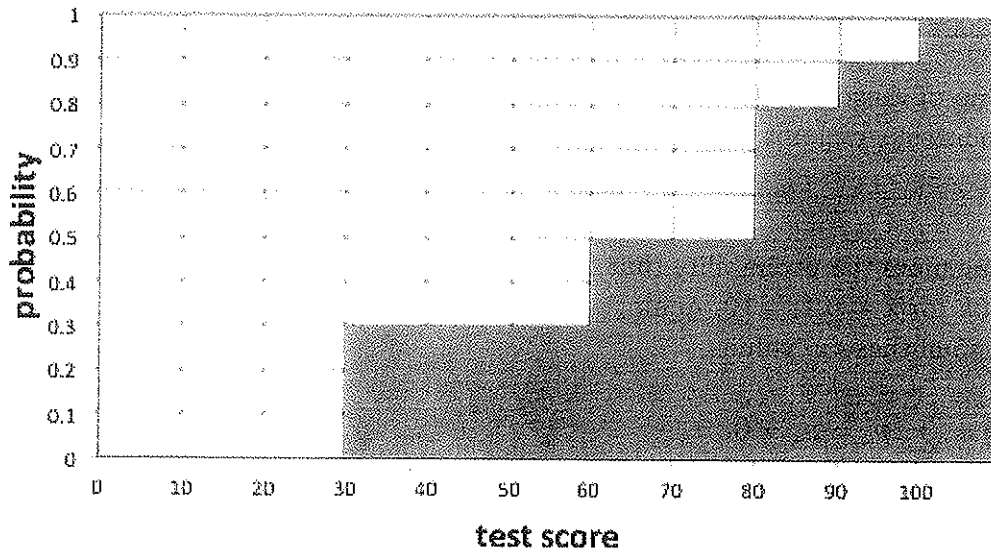
1.5 Some Common Discrete Distributions

A random variable is a theoretical representation of a physical or experimental process we wish to study. Formally, it is a function defined over a *sample space* of possible outcomes. For our simple coin tossing experiment, where we flip a fair coin once and observe the outcome, our sample space consists of the two outcomes, H or T. When tossing two fair coins sequentially, our sample space consists of the four outcomes HH, HT, TH or TT.

Let us fix a sample space of n tosses of a fair coin. Experimentally, we may be interested in studying the number of "heads" observed after tossing the coin n times. Or we could be interested in studying the number of tosses needed to first observe "heads". Or we could be interested in studying how likely a certain sequence of "heads" and "tails" is to be observed. Each of these experiments are defined on the same sample space (the events generated by n tosses of a fair coin), yet each strive to quantify different things. Consequently, each experiment should be associated with a different random variable.

The Binomial Distribution

Let X_n denote the random variable that counts the number of times we observe "heads" when flipping a fair coin n times. Clearly, X can take on any integer value from 0 to n, corresponding to the experimental outcome of observing 0 to n "heads". How likely is any particular outcome of this random variable? Notice that we do not care about the order of the observations here, so that if $n = 3$, the outcome THH is equivalent to the outcomes HTH and HHT. Each of these outcomes contains two "heads".

The likelihood of any particular outcome is what is represented by the probability mass function (PMF) of the random variable. Suppose $n = 2$. Then we see that the PMF of X_2 is given by:

- $Pr(X_2 = 0) = 1/4$

- $\Pr(X_2 = 1) = 1/2$
- $\Pr(X_2 = 2) = 1/4$

We say that X_2 is a **binomial** random variable with parameters 2 (the number of times we flip the fair coin) and 1/2 (the probability that we observe heads after a single flip of the coin). We can write $X_2 \sim \mathrm{Bin}(2, 1/2)$.

Just as we did with Bernoulli random variables, we can think of our coin tossing experiment a bit more abstractly. Specifically, we can think of observing "heads" as a success and observing "tails" as a failure. This abstraction will help us generalize our coin tossing procedure to more general experiments.

If X is a binomial random variable associated to n independent trials, each with a success probability p, then the probability mass function of X is:

$$\Pr(X = k) = \frac{n!}{k!(n - k)!} \cdot p^k(1 - p)^{n-k},$$

where k is any integer from 0 to n. Recall that the *factorial* notation $n!$ denotes the product of the first n positive integers: $n! = 1 \cdot 2 \cdot 3 \cdots (n\text{-}1) \cdot n$, and that we observe the convention $0! = 1$.

For our coin tossing experiment, the probability of success - that is, the probability of observing "heads" - was the same as the probability of failure, observing "tails". In general, we may be interested in processes that have different probabilities of success and failure.

For example, suppose that we know that 5% of all light bulbs produced by a particular manufacturer are defective. If we buy a package of 6 light bulbs and want to calculate the probability that at least one is defective, we can do so by identifying this experiment with a binomial random variable. Here, we can think of observing a defective bulb as a "success" and observing a functional bulb as a "failure". Then our experiment is given by the random variable $X_6 \sim \mathrm{Bin}(6, 1/20)$, since we will observe 6 bulbs in total and each has a probability of $5/100 = 1/20$ of being defective.

In general, we can think of observing n independent experimental trials and counting the number of "successes" that we witness. The probability distribution we associate with this setup is the **binomial** random variable with parameters n and p, where p is the probability of "success." We can denote this distributional relationship to a random variable X by $X \sim \mathrm{Bin}(n, p)$.

The Geometric Distribution

Now consider a slightly different experiment where we wish to flip our fair coin repeatedly until we first observe "heads". Since we can first observe heads on the first flip, the second flip, the third flip, or on any subsequent flip, we see that the possible values our random variable can take are 1, 2, 3,....

Of course, we can consider a more abstract experiment where we observe a sequence of trials until we first observe a success, where the probability of success is p. If we let X denote such a random variable, then we say that X is a **geometric** random variable with parameter p. We can denote this particular random variable by $X \sim \mathrm{Geo}(p)$.

Letting S denote the outcome of "success" and F denote the outcome of "failure", we can summarize the possible outcomes of a geometric experiment and their likelihoods (the probability mass function) in the following table. Here, we write p for the probability of success and q for the probability of failure.

Experimental Outcome	Value of the Random Variable, $X = x$	Probability
S	$x = 1$	p
FS	$x = 2$	$q \cdot p$
FFS	$x = 3$	$q^2 \cdot p$
FFFS	$x = 4$	$q^3 \cdot p$
FFFFS	$x = 5$	$q^4 \cdot p$
...	...	...

When flipping a fair coin, we see that $X \sim \text{Geo}(1/2)$, so that our PDF takes the particularly simple form $\Pr(X = k) = (1/2)^k$ for any positive integer k.

The Discrete Uniform Distribution

Now consider a coin tossing experiment of flipping a fair coin n times and observing the sequence of "heads" and "tails". Because each outcome of a single flip of the coin is equally likely, and because the outcome of a single flip does not affect the outcome of another flip, we see that the likelihood of observing any particular sequence of "heads" and "tails" will always be the same. Notice that for $n = 2$ or 6, we have already encountered this random variable (see Section 1.01 and Sections 1.02 - 1.04 respectively).

We say that a random variable X has a **discrete uniform** distribution on n points if X can assume any one of n values, each with equal probability. Evidently then, if X takes integer values from 1 to n, we find that the PMF of X must be $\Pr(X = k) = 1/n$, for any integer k between 1 and n.

1.6 Expected Value

For an experiment or general random process, the outcomes are never fixed. We may replicate the experiment and generally expect to observe many different outcomes. Of course, in most reasonable circumstances we will expect these observed differences in the outcomes to collect with some level of concentration about some central value. One central value of fundamental importance is the *expected value*.

The **expected value** or **expectation** (also called the **mean**) of a random variable X is the weighted average of the possible values of X, weighted by their corresponding probabilities. Informally, the expectation of a random variable X is the average value that we would expect to see after repeated observation of the random process. Put another way, the expectation is the long-term average of the realized values of a random variable after repeated observation of the random variable.

Expected Value of a Discrete Random Variable

The expected value, $\mathbb{E}(X)$, of a discrete random variable X is the weighted average of the possible values of X where each possible value of X is weighted by its corresponding probability:

$$\mathbb{E}(X) = \sum_{k=1}^{N} x_k \Pr(x = x_k)$$

where N is the total number of possible values of X.

- Do not confuse the *expected* value with the *average* value of a set of observations: they are two different but related quantities. The average value of a random variable X would be just the ordinary average of the possible values of X; that is, no possible value of X receives any special weight. Naturally, this ordinary average is given by $\frac{1}{N}\sum_{k=1}^{N} x_k$. The expected value of X is a *weighted* average, where certain values get more or less weight depending on how likely or not they are to be observed. A true average value is calculated only when all weights (so all probabilities) are the same.
- The definition of expected value requires numerical values for the x_k. So if the outcome for an experiment is something qualitative, such as "heads" or "tails", we could calculate the expected value if we assign heads and tails numerical values (0 and 1, for example).

Example: Test Scores

Recall the test score example from Sections 1.03 and 1.04. We supposed that in a class of 10 people the grades on a test are given by 30, 30, 30, 60, 60, 80, 80, 80, 90, 100. A test is drawn from the collection at random and the score X is observed. What is the expected value of the random variable X?

The expected value of the random variable is given by the weighted average of its values:

$$\mathbb{E}(X) = \sum_{k=1}^{N} x_k \Pr(X = x_k)$$

$$= 30\frac{3}{10} + 60\frac{2}{10} + 80\frac{3}{10} + 90\frac{1}{10} + 100\frac{1}{10}$$

$$= 9 + 12 + 24 + 9 + 10$$

$$= 64$$

Notice that 64 is not actually a possible value for the random variable X. Nevertheless, this expectation makes sense if we remember that what we have really calculated is the long-term average of repeatedly drawing a test score from this collection. If we drew a test score at random from this collection 100 times (remembering to replace the selected test each time so that we never alter our collection of tests) and then averaged all the observed outcomes, this average value would be very near the expected value of 64.

Expectation as a Measure of the Center of a Distribution

Another informal way to think of the expectation of a random variable is to notice that it gives a measure of the center of the associated distribution. For our test score example, the PMF of the randomly selected test score X is shown below.

Probability Distribution for X
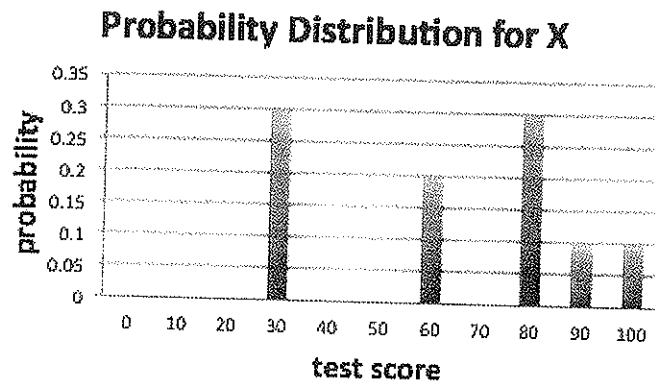
Notice that the expected value of our randomly selected test score, $\mathbb{E}(X) = 64$, lies near the "center" of the PMF. There are many different ways to quantify the "center of a distribution" - for example, computing the 50th percentile of the possible outcomes - but for our purposes we will concentrate our attention on the expected value.

1.7 Variance and Standard Deviation

Another important quantity related to a given random variable is its variance. The **variance** is a numerical description of the spread, or the *dispersion*, of the random variable. That is, the variance of a random variable X is a measure of how spread out the values of X are, given how likely each value is to be observed.

Variance and Standard Deviation of a Discrete Random Variable

The variance, Var(X), of a discrete random variable X is

$$\text{Var}(X) = \sum_{k=1}^{N} \left(x_k - \mathbb{E}(X) \right)^2 \Pr(X = x_k)$$

where N is the total number of possible values of X.

The **standard deviation**, σ, is the positive square root of the variance:

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

Observe that the variance of a random variable is always nonnegative (since probabilities are nonnegative, and the square of a number is also nonnegative).

Observe also that much like the expectation of a random variable X, the variance (or standard deviation) is a weighted average of an expression of observable and calculable values. More precisely, notice that

$$\text{Var}(X) = \mathbb{E}\left([X - \mathbb{E}(X)]^2 \right).$$

Example: Test Scores

Using the test scores example of the previous sections, calculate the variance and standard deviation of the random variable X associated to randomly selecting a single exam.

Solution

The variance of the random variable X is given by

$$\text{Var}(X) = \sum_{k=1}^{N} (x_k - \mathbb{E}(X))^2 \Pr(X = x_k)$$

$$= (30 - 64)^2 \frac{3}{10} + (60 - 64)^2 \frac{2}{10} + (80 - 64)^2 \frac{3}{10} + (90 - 64)^2 \frac{1}{10} + (100 - 64)^2 \frac{1}{10}$$

$$= 624$$

The standard deviation of X is then

$$\sigma(X) = \sqrt{624} \approx 24.979992$$

Interpretation of the Standard Deviation

For most "nice" random variables, i.e. ones that are not too wildly distributed, the standard deviation has a convenient informal interpretation. Consider the intervals

$$S_m = [\mathbb{E}(X) - m\sigma(X),\ \mathbb{E}(X) + m\sigma(X)],$$

for some positive integer m. As we increase the value of m, these intervals will contain more of the possible values of the random variable X.

A good rule of thumb is that for "nicely distributed" random variables, all of the most likely possible values of the random variable will be contained in the interval S_3. Another way to say this is that, for discrete random variables, most of the PMF will live on the interval S_3. We will see in the next chapter that a similar interpretation holds for continuous random variables.

1.8 Chapter 1 Summary

Our treatment of discrete random variables has been brief, but the concepts we have introduced are fundamental to any random process. These fundamentals will be explored again in the next chapter when we apply them to continuous random variables. We will see that many similarities exist between the discrete and continuous cases, but we will also notice many important differences between the two as well.

We summarize some of the important concepts that were introduced in Chapter 1.

The PMF and the CDF

The **probability mass function** (PMF) of a random variable X is the function that assigns probabilities to the possible outcomes of X. We write

$$\boxed{\Pr(X = x_k)}$$

to denote this function of the possible values x_k of X.

The **cumulative distribution function** (CDF) of a random variable X is the function that accumulates the probabilities from a specified value. We define the CDF to be $F(x) = \Pr(X \le x)$ and note that the CDF is intimately related to the PMF via our identity for the probability of disjoint events; i.e., the CDF is given by a sum over values of the PMF.

$$\boxed{F(x_n) = \sum_{k=1}^{n} \Pr(X = x_k)}$$

Expected Value, Variance and Standard Deviation

The concepts of **expectation**, **variance** and **standard deviation** are crucial and will be revisited again when we explore continuous random variables. Students should know that

$$\boxed{\begin{aligned} \mathbb{E}(X) &= \sum_{k=1}^{N} x_k \Pr(X = x_k) \\ \mathrm{Var}(X) &= \sum_{k=1}^{N} (x_k - \mathbb{E}(X))^2 \cdot \Pr(X = x_k) \\ \sigma(X) &= \sqrt{\mathrm{Var}(X)} \end{aligned}}$$

The expectation represents the "center" of a random variable, an expected value of an experiment, or the average of outcomes of an experiment repeated many times. The variance and standard deviation of a random variable are numerical measures of the spread, or dispersion, of the distribution of the random variable.

Continuous Random Variables

2.1 The Cumulative Distribution Function (Continuous Case)

In the previous chapter, we defined random variables in general, but focused only on **discrete** random variables. In this chapter, we properly treat **continuous** random variables.

If for example X is the height of a randomly selected person in British Columbia, or X is tomorrow's low temperature at Vancouver International Airport, then X is a continuously varying quantity.

We previously defined a **continuous random variable** to be one where the values the random variable can assume are given by a *continuum* of values. For example, we can define a continuous random variable that can take on any value in the interval [1,2].

To make this definition more precise, we recall the definition from Section 1.4 of a cumulative distribution function (CDF) that was given for *any* random variable.

The Cumulative Distribution Function

The **cumulative distribution function** for *any* random variable X, denoted by $F(x)$, is the probability that X assumes a value less than or equal to x:

$$F(x) = \Pr(X \leq x)$$

The cumulative distribution function has the following properties:

- $0 \leq F(x) \leq 1$ for all values of x
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to +\infty} F(x) = 1$
- $F(x)$ is a nondecreasing function of x

This definition is independent of the type of random variable to which we are referring. From this, we can define a **continuous random variable** to be any random variable X whose CDF is a *continuous* function. Notice that this is in contrast to the case of discrete random variables where the corresponding CDF is always a discontinuous step-function.
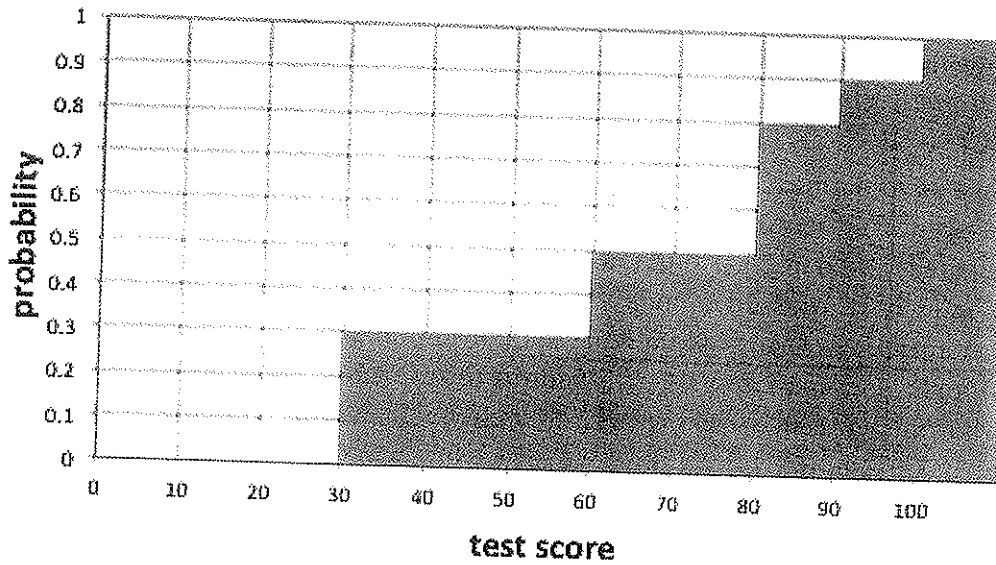

Continuous Random Variable

A **continuous random variable** is one that has a *continuous* cumulative distribution function.


Comparison to the Discrete Case

Recall the cumulative distribution function we had for the test scores example in the previous chapter. The graph of the cumulative distribution function is given below.

Cumulative Distribution for X

Observe that the graph of this function "increases in steps" at 30, 60, 80, 90, and 100. CDFs for discrete random variables are *always* step functions. A discrete random variable cannot assume a continuum of values; thus, its CDF can only increase at a finite or countably infinite set of points.

For another example, consider the function whose graph is given below.
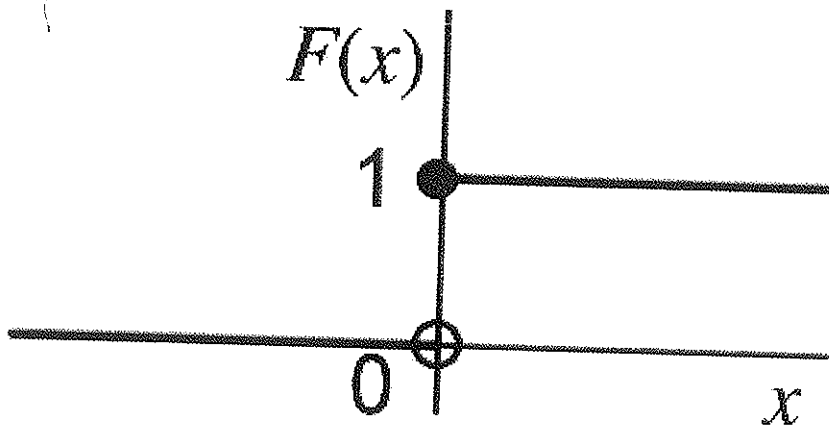
This function cannot represent a CDF for a *continuous* random variable because the function F is not continuous for all values of x. However, F could represent a cumulative distribution function for a *discrete* random variable since it increases from 0 to 1 in a finite number of steps.

Example: Maximum Outdoor Air Temperature

The maximum outdoor air temperature in downtown Vancouver on any given day in January can be expressed as a continuous random variable X. A reasonable CDF for this random variable is given by the function

$$F(x) = \frac{1}{1 + e^{-kx}}, \quad \text{for some } k > 0$$

For a particular k, we have graphed this cumulative distribution function in the plot below.
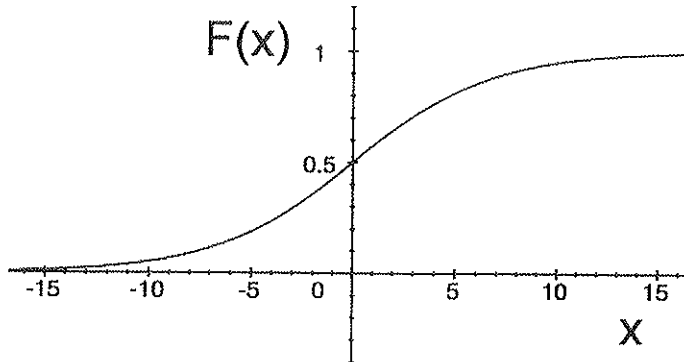
In the above plot, note that the horizontal x-axis gives possible values of the maximum outdoor air temperature in downtown Vancouver on any day in January, and that the vertical probability-axis gives values between 0 and 1. The value of the cumulative distribution function $F(x)$ gives the probability that the maximum outdoor air temperature is no greater than x.

We can easily see that this function satisfies the basic properties of a CDF. Clearly, $F(x) \geq 0$ for all possible temperatures x. Also, $F(x) \leq 1$ for all x since the denominator in the definition of $F(x)$ is always larger than the numerator. Since $k > 0$, we calculate

$$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} \frac{1}{1 + e^{-kx}} = 0.$$

Likewise,

$$\lim_{x \to +\infty} F(x) = \lim_{x \to +\infty} \frac{1}{1 + e^{-kx}} = 1.$$

To check that F is nondecreasing, we note that since F is everywhere differentiable, it suffices to show that the derivative of F is nonnegative. A quick calculation yields

$$\frac{d}{dx} F(x) = \frac{ke^{-kx}}{(1 + e^{-kx})^2}$$

which is certainly never negative. Thus we see explicitly that this function $F(x)$ satisfies all the basic properties that a CDF should.

2.2 The Probability Density Function

An Important Distinction Between Continuous and Discrete Random Variables

What is $\Pr(X = x)$? The answer clearly depends on the distribution of the random variable X. For discrete random variables, we have already seen that if x is a possible value that X can assume, then $\Pr(X = x)$ is some positive number. But is this still true if X is a continuous random variable?

In the context of our maximum outdoor air temperature example from the previous section, we may ask *what is the probability that the maximum outdoor air temperature in downtown Vancouver on any given day in January is **exactly** 0°C?* Since our measurements of the air temperature are never exact, this probability should be *zero*. If we had instead asked for the probability that the maximum outdoor air temperature was within 0.005° of 0°C, then we would have arrived at a nonzero probability. All practical measurements of continuous data are always approximate. They may be very precise, but they can never be truly *exact*. Hence, we cannot expect to measure the likelihood of an exact outcome, only an approximate one.

In general, for any continuous random variable X, we will always have $\Pr(X = x) = 0$. We can prove this fact directly by appealing to our basic results about combining probabilities of disjoint events.

Suppose we choose any interval $[x, x + \Delta x]$. The probability that the continuous random variable X lies inside of this interval is

$$\Pr(x \leq X \leq x + \Delta x).$$

Using our identity for probabilities of disjoint events, we can write this as the difference

$$\Pr(X \leq x + \Delta x) - \Pr(X \leq x).$$

If we take the limit as Δx goes to zero, we obtain

$$\lim_{\Delta x \to 0} \Pr(x \leq X \leq x + \Delta x) = \lim_{\Delta x \to 0} \left[\Pr(X \leq x + \Delta x) - \Pr(X \leq x) \right]$$
$$= \Pr(X \leq x) - \Pr(X \leq x)$$
$$= 0$$

Notice that the crucial step in this argument is the evaluation of the limit in the second to last line. Since X is a *continuous* random variable, its CDF $F(x)$ is a continuous function; thus, we are allowed to pass the limit through to the argument of the function $F(x) = \Pr(X \leq x)$. Notice that if X were a discrete random variable, this evaluation would not be possible in general since its CDF would not be continuous.

This gives a direct proof of the fact that $\Pr(X = x) = 0$ for any continuous random variable X. We will see that an even simpler proof will come for free for most continuous random variables via the Fundamental Theorem of Calculus. In order to do this however, we need to relate these probabilities to an integration of some appropriate function. It turns out that this function plays a vital role in describing the distribution of a continuous random variable and will be extremely useful for performing calculations.

The Probability Density Function

The "appropriate function" referred to above is called the **probability density function** (PDF). It can be defined for most continuous random variables, and is extremely useful for calculating probabilities of events associated to a continuous random variable.

The Probability Density Function

Let $F(x)$ be the cumulative distribution function for a continuous random variable X. The **probability density function** (PDF) of X is given by

$$f(x) = \frac{dF(x)}{dx},$$

wherever the derivative exists.

In short, the PDF of a continuous random variable is the derivative of its CDF. Using the Fundamental Theorem of Calculus, we see that the CDF $F(x)$ of a continuous random variable X may be expressed in terms of its PDF:

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

where f denotes the PDF of X.

Properties of the PDF

This formulation of the PDF via the Fundamental Theorem of Calculus allows us to derive the following properties.

Properties of the Probability Density Function

If $f(x)$ is a **probability density function** for a continuous random variable X, then

1. $F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(t)dt$

2. $f(x) \ge 0$ for any value of x

3. $\int_{-\infty}^{\infty} f(t)dt = 1$

The first property, as we have already seen, is just an application of the Fundamental Theorem of Calculus and relates the CDF of a continuous random variable to its PDF.

The second property states that for a function to be a PDF, it must be nonnegative. This makes intuitive sense since probabilities are always nonnegative numbers. More precisely, we already know that the CDF $F(x)$ is a nondecreasing function of x. Thus, its derivative $f(x)$ is nonnegative.

The third property states that the area between the function and the x-axis must be 1, or that all probabilities must integrate to 1. This must be true since $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$; thus Property 3 follows from the Fundamental Theorem of Calculus.

The PDF gives us a helpful geometrical interpretation of the probability of an event: the probability that a continuous random variable X is less than some value x_0, is equal to the area under the PDF $f(x)$ on the interval $(-\infty, x_0]$, as demonstrated in the following graph.
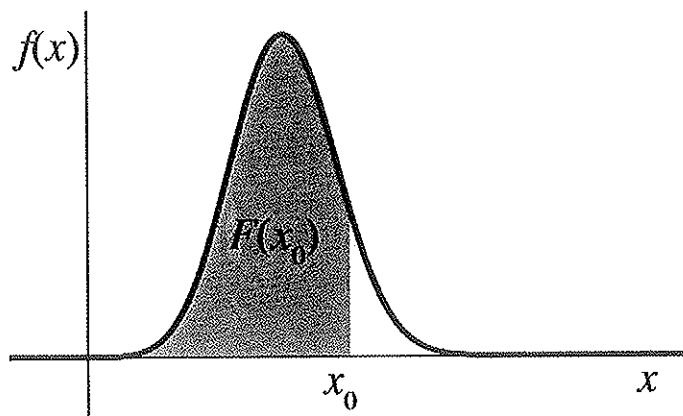
Similarly, we have $\Pr(a \leq x \leq b) = \int_a^b f(x)dx$.

Now that we can interpret probabilities as integrals, it is clear that for a continuous random variable X, we will always have $\Pr(X = x) = 0$. This is simply because the area under a single point of a curve is always zero. In other words, if X is a continuous random variable, the probability that X is equal to a particular value will always be zero. We again note that this is an important difference between continuous and discrete random variables.

The PDF of a continuous random variable plays a similar role as the PMF does for discrete random variables. In particular, they are both used to compute probabilities of events associated to a random variable. However, as the previous paragraph shows, PDFs and PMFs are different objects, just as continuous and discrete random variables are different concepts.

Example

Let $f(x) = k(3x^2 + 1)$ for $0 \leq x \leq 2$, and $f(x) = 0$ elsewhere.

1. Find the value of k that makes the given function a PDF.
2. Let X be a continuous random variable whose PDF is $f(x)$. Compute the probability that X is between 1 and 2.
3. Find the cumulative distribution function of X.
4. Find the probability that X is *exactly* equal to 1.

Solution

Part 1)

$$1 = \int_{-\infty}^{\infty} f(x)dx$$
$$= \int_0^2 k(3x^2 + 1)dx$$
$$= k\left(\frac{3x^3}{3} + x\right)\Big|_0^2 dx$$
$$= k(10)$$

Therefore, $k = 1/10$.

Notice that $f(x) \geq 0$ for all x. Also notice that we can rewrite this PDF as a piecewise function:

$$f(x) = \begin{cases} \frac{1}{10}(3x^2 + 1) & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Part 2)

Using our value of k from Part 1:

$$\Pr(1 \le X \le 2) = \int_1^2 \frac{3x^2 + 1}{10} dx = \frac{x^3 + x}{10} \Big|_1^2 = 1 - 2/10 = 4/5$$

Therefore, $\Pr(1 \le X \le 2)$ is 4/5.

Part 3)

Using the Fundamental Theorem of Calculus, the CDF of X at x in [0,2] is

$$\begin{aligned} \Pr(X \le x) = F(x) &= \int_{-\infty}^x f(t)dt \\ &= \int_0^x \frac{1}{10}(3t^2 + 1)dt \\ &= \frac{1}{10}(t^3 + t)\Big|_0^x \\ &= \frac{1}{10}(x^3 + x), \text{ for } 0 \le x \le 2 \end{aligned}$$

A similar calculation easily verifies that $F(x) = 0$ for all $x < 0$ and that $F(x) = 1$ for all $x > 2$.

Part 4)

Since X is a continuous random variable, we immediately know that the probability that it equals any one particular value must be zero. More directly, we compute

$$\Pr(X = 1) = \int_1^1 f(t)dt = 0$$

An Important Subtlety

There is an important subtlety in the definition of the PDF of a continuous random variable. Notice that the PDF of a continuous random variable X can only be defined when the cumulative distribution function of X is *differentiable*.

As a first example, consider the experiment of randomly choosing a real number from the interval [0,1]. Let X denote the outcome of this experiment. Since the likelihood of picking a number in a given subinterval of [0,1] is proportional to the length of that subinterval, we see that the CDF $F(x)$ is given by

$$\Pr(X \le x) = F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

This function is differentiable everywhere *except* at the points $x = 0$ and $x = 1$. So the PDF of X is defined at all points except for these two:

$$\frac{dF(x)}{dx} = f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{if } x < 0 \text{ or } x > 1 \end{cases}$$

Nevertheless, it can still make sense to define the PDF at the points where the CDF fails to be differentiable. We know that the integral over a single point is always zero, so we can always change the value of our PDF at any particular point (or at any finite set of points) without changing the probabilities of events associated to our random variable. Thus, we could define

$$\frac{dF(x)}{dx} := f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

or

$$\frac{dF(x)}{dx} := f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Both of these functions are also PDFs of the continuous random variable X. These two formulations have the advantage of being defined for all real numbers.

Not All Continuous Random Variables Have PDFs

We can sometimes encounter continuous random variables that simply do not have a meaningful PDF at all. The simplest such example is given by a distribution function called the *Cantor staircase*.

The *Cantor set* is defined recursively as follows:

* Start with the interval $[0,1)$.
* Delete the middle third of this interval. You are now left with two subintervals $[0,1/3)$ and $[2/3,1)$.
* Delete the middle third of each of these remaining subintervals. Now we have four new subintervals: $[0,1/9)$, $[2/9,3/9)$, $[6/9,7/9)$, and $[8/9,1)$.
* Repeat this middle third deletion for the new subintervals. Continue indefinitely.

If we take this process to the limit, the set that remains is called the *Cantor set*. It is extremely sparse in $[0,1)$, yet still contains about as many points as the entire interval itself. In particular, notice that every point of the form $x = 1 - 3^{-k}$ is in the Cantor set for every $k > 0$.

We can define a Cantor random variable to have the cumulative distribution function that increases on the Cantor set and remains constant off of this set. We define this function as follows:

* Let $F(x)$ be the CDF of our Cantor random variable X. Define $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$.
* Define $F(x) = 1/2$ on $[1/3,2/3)$, i.e. on the first middle third deleted in the construction of the Cantor set.
* Define $F(x) = 1/4$ on $[1/9,2/9)$ and $F(x) = 3/4$ on $[7/9,8/9)$.
* Define $F(x) = 1/8, 3/8, 5/8,$ and $7/8$ on the deleted middle thirds from the third step in our Cantor set construction.
* Continue indefinitely.

After a limiting argument and some technicalities with defining $F(x)$ on the Cantor set itself, this procedure defines a *continuous* function that begins at 0 and increases to 1. However, since this function is *constant* except on the Cantor set, we see that its derivative off of the Cantor set must be identically *zero*. On the Cantor set the function is not differentiable and so has no natural PDF.

What we see is that, for a Cantor random variable, we cannot make any sensible definition for the PDF. It is either identically zero or not defined.

This is an interesting example of how identifying a random variable with its PDF can lead us astray. Thankfully, for our purposes, we will never need to consider continuous random variables that do not have PDFs defined everywhere (except possibly at finitely many points).

2.3 Some Common Continuous Distributions

Let us consider some common continuous random variables that often arise in practice. We should stress that this is indeed a very small sample of common continuous distributions.

The Beta Distribution

Suppose the proportion p of restaurants that make a profit in their first year of operation is given by a certain *beta* random variable X, with probability density function:

$$f(p) = \begin{cases} 12p(1-p)^2 & \text{if } 0 \leq p \leq 1, \\ 0 & \text{elsewhere.} \end{cases}$$

What is the probability that more than half of the restaurants will make a profit during their first year of operation? To answer this question, we calculate the probability as an area under the PDF curve as follows:

$$\begin{aligned} \Pr(0.5 \leq X \leq 1) &= \int_{0.5}^{1} f(p)dp \\ &= \int_{0.5}^{1} 12p(1-p)^2 dp \\ &= \int_{0.5}^{1} \left(12p - 24p^2 + 12p^3\right) dp \\ &= 6p^2 - 8p^3 + 3p^4 \Big|_{0.5}^{1} \\ &= (6 - 8 + 3) - (1.5 - 1 + 0.1875) \\ &= 0.3125 \end{aligned}$$

Therefore, $\Pr(0.5 \leq P \leq 1) = 0.3125$.

The example above is a particular case of a beta random variable. In general, a **beta random variable** has the generic PDF:

$$f(x) = \begin{cases} kx^{a-1}(1-x)^{b-1} & \text{if } 0 \leq x \leq 1, \\ 0 & \text{elsewhere} \end{cases}$$

where the constants a and b are greater than zero, and the constant k is chosen so that the density f integrates to 1.

We see that our previous example was a beta random variable given by the above density with $a = 2$ and $b = 3$. Let us find the associated cumulative distribution function $F(p)$ for this random variable. We compute:

$$F(p) = \int_{-\infty}^{p} f(t)dt$$

$$= \int_{0}^{p} 12t(1-t)^2 dt$$

$$= 12 \int_{0}^{p} (t - 2t^2 + t^3)dt$$

$$= 12\left(\frac{1}{2}t^2 - \frac{2}{3}t^3 + \frac{1}{4}t^4\right)\Big|_{0}^{p}$$

$$= p^2(6 - 8p + 3p^2),$$

valid for $0 \le p \le 1$.

The Exponential Distribution

The lifespan of a lightbulb can be modeled by a continuous random variable since lifespan - i.e. *time* - is a continuous quantity. A reasonable distribution for this random variable is what is known as an *exponential distribution*.

A random variable Y has an **exponential distribution with parameter $\beta > 0$** if its PDF is given by

$$f(y) = \begin{cases} \frac{1}{\beta}e^{-y/\beta} & \text{if } 0 \le y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

Suppose that the lifespan (in months) of lightbulbs manufactured at a certain facility can be modeled by an exponential random variable Y with parameter $\beta = 4$. What is the probability that a particular lightbulb lasts at least a year? Again, we can calculate this probability by evaluating an integral. Since there are 12 months in one year, we calculate

$$\Pr(Y \ge 12) = \int_{12}^{\infty} f(y)dy$$

$$= \int_{12}^{\infty} \frac{1}{4}e^{-y/4}dy$$

$$= -e^{-y/4}\Big|_{12}^{\infty}$$

$$= 0 - (-e^{-3})$$

$$\approx 0.04979$$

Thus we can see that it is highly likely we would need to replace a lightbulb produced from this facility within one year of manufacture.

The Continuous Uniform Distribution

Our third example of a common continuous random variable is one that we have already encountered. Consider the experiment of randomly choosing a real number from the interval [a,b]. Letting X denote this random outcome, we say that X has a *continuous uniform* distribution on [a,b] if the probability that we choose a value in some subinterval of [a,b] is given by the relative size of that subinterval in [a,b]. More explicitly, we have the following:

A random variable X has an **continuous uniform** distribution on [a,b] if its PDF is constant on [a,b]; i.e. its PDF is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$$

The continuous uniform distribution has a particularly simple representation, just as its discrete counterpart does. Nevertheless, this random variable has great practical and theoretical utility. We will explore this distribution in more detail in the following example and in the exercises.

A Geometric Problem

Consider the square in the xy-plane bounded by the lines $x = 0$, $x = 1$, $y = 0$ and $y = 1$. Now consider a vertical line with equation $x = b$, where $0 \le b \le 1$ is fixed. Note that this line will intersect the unit square just defined.

Suppose we select a point inside this square, uniformly at random. If we let X be the x-coordinate of this random point, what is the probability that X is in the interval $[0, b]$?

An illustration of our problem is given in the figure below. Graphically, we are trying to find the probability that a randomly selected point inside the square lies to the *left* of the red line.
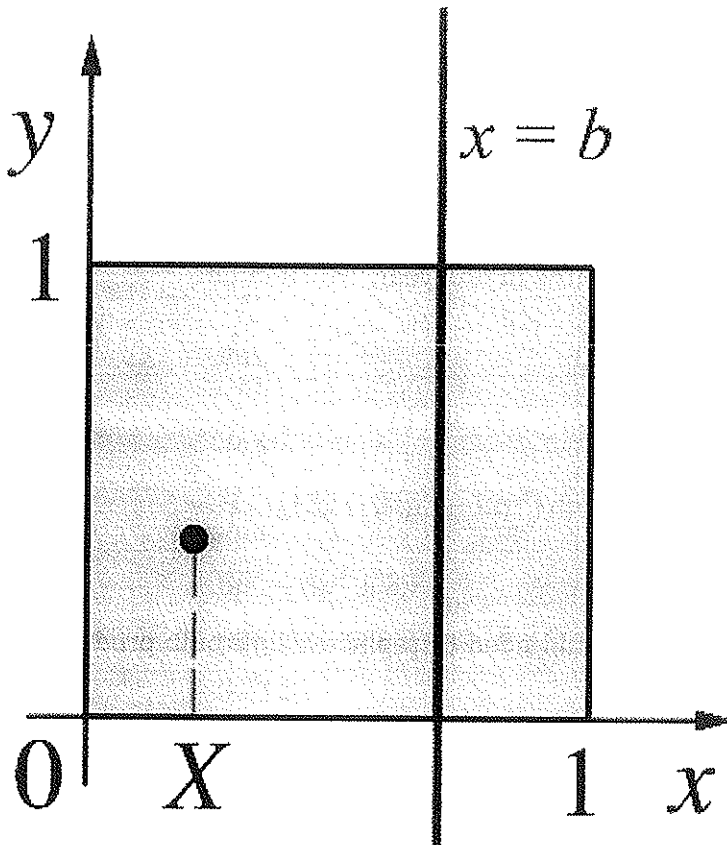
The region to the left of the red line is a rectangle with area equal to b. The probability that our random point lies inside this rectangle is proportional to the area of that rectangle, since the larger the area of the rectangle, the larger the probability is that the point is inside of it.

- If the probability that the point is between 0 and b were equal to 0.5, then the red line would have to divide the square into two equal halves: so $b = 0.5$.
- If the probability that the point is between 0 and b were equal to 0.25, then the red line would have to divide the square at 1/4: so $b = 0.25$.

- If the probability that the point is between 0 and b were equal to 1, then the red line would have to lie on the rightmost edge of the square itself: so $b = 1$

In general, we see that we should have $\Pr(0 \leq X \leq b) = b$.

Notice that this result matches with the definition of our random variable X. Since we want to select a random point *uniformly at random* from the unit square, the random variable X giving the x-coordinate of this random point should be a *continuous uniform* random variable on the interval $[0,1]$. Thus, the PDF of X is simply $f(x) = 1$, where $0 \leq x \leq 1$.

Therefore, $\Pr(0 \leq X \leq b) = \int_0^b dx = b$, which agrees with the answer we derived using purely geometric considerations.

2.4 The Normal Distribution

The most important probability distribution in all of science and mathematics is the *normal distribution*.

The Normal Distribution

The random variable X has a **normal distribution with mean parameter μ and variance parameter $\sigma^2 > 0$** if and only if its PDF is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

To express this distributional relationship on X, we commonly write $X \sim \text{Normal}(\mu, \sigma^2)$.

This PDF is the classic "bell curve" shape associated with so many experiments and natural phenomena. The parameter μ gives the mean of the distribution (the center of the bell curve) while the σ^2 parameter gives the variance (the horizontal spread of the bell curve). The first of these facts is a simple exercise in integration (see the exercises), while the second requires a bit more ingenuity.

Recall that the standard deviation of a random variable is defined to be the positive square root of its variance. Thus, a normal random variable has standard deviation σ.

This random variable enjoys many analytical properties that make it a desirable object to work with theoretically. For example, the normal density is symmetric about its mean μ. This means that, among other things, exactly half of the area under the PDF lies to the right of the mean, and the other half of the area lies to the left of the mean. More generally, we have the following important fact.

Symmetry of Probabilities for a Normal Distribution

If X has a normal distribution with mean μ and variance σ^2, and if x is any real number, then

$$\Pr(X \leq \mu - x) = \Pr(X \geq \mu + x).$$

However, the PDF of a normal distribution is not convenient for calculating probabilities directly. In fact, it can be shown that no closed form exists for the cumulative distribution function of a normal random variable. Thus, we must rely on tables of values to calculate probabilities for events associated to a normal random variable. (The values in these tables are calculated using numerical techniques of integration.)

A particularly useful version of the normal distribution is the *standard normal* distribution, where the mean parameter is 0 and the variance parameter is 1.

The Standard Normal Distribution

The random variable Z has a **standard normal distribution** if its distribution is normal with mean 0 and variance 1. The PDF of Z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty.$$

For a particular value x of a normal random variable X, the distance from x to the mean μ of X expressed in units of standard deviation σ is

$$z = \frac{x - \mu}{\sigma}.$$

Since we have subtracted off the mean (the center of the distribution) and factored out the standard deviation (the horizontal spread), this new value z is not only a rescaled version of x, but is also a realization of a *standard* normal random variable.

In this way, we can standardize any value from a generic normal distribution, transforming it into one from a standard normal distribution. Thus we reduce the problem of calculating probabilities for an event from a normal random variable to calculating probabilities for an event from a *standard* normal random variable.

Standardizing a Normal Random Variable

Let X have a normal distribution with mean μ and variance σ^2. Then the new random variable

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution.

Calculating Probabilities of Events Using a Standard Normal Distribution

Suppose that the test scores for first-year integral calculus final exams are normally distributed with mean 70 and standard deviation 14. Given that $\Pr(Z \le 0.36) = 0.64$ and $\Pr(Z \le 1.43) = 0.92$ for a standard normal random variable Z, what percentage of final exam scores lie between 75 and 90?

If we let X denote the score of a randomly selected final exam, then we know that X has a normal distribution with parameters $\mu = 70$ and $\sigma = 14$. To find the percentage of final exam scores that lie between 75 and 90, we need to use the information about the probabilities of a *standard* normal random variable. Thus we must standardize X using the technique above.

For our particular problem, we wish to compute

$$\Pr(75 \le X \le 90).$$

We proceed by standardizing the random variable X as well as the particular x values of interest. Thus, since X has mean 70 and standard deviation 14, we write

$$\Pr(75 \le X \le 90) = \Pr\left(\frac{75 - 70}{14} \le \frac{X - 70}{14} \le \frac{90 - 70}{14}\right).$$

Now we have standardized our normal random variable so that

$$\frac{X - 70}{14} = Z,$$

where $Z \sim \text{Normal}(0,1)$.

Simplifying the numerical expressions from above, we deduce that we must calculate

$\Pr(0.36 \leq Z \leq 1.43)$.

Now we can use the information we were given, namely that $\Pr(Z \leq 0.36) = 0.64$ and $\Pr(Z \leq 1.43) = 0.92$. Using these values, we find

$$\begin{aligned}
\Pr(75 \leq X \leq 90) &= \Pr(0.36 \leq Z \leq 1.43) \\
&= \Pr(Z \leq 1.43) - \Pr(Z \leq 0.36) \\
&= 0.92 - 0.64 \\
&= 0.28.
\end{aligned}$$

Therefore the percentage of first-year integral calculus final exam scores between 75 and 90 is 28%.

Now suppose we wish to find the percentage of final exam scores larger than 90, as well as the percentage of final exam scores less than 65. To find the percentage of final exam scores larger than 90, we use our knowledge about probabilities of disjoint events:

$$\begin{aligned}
\Pr(X > 90) &= 1 - \Pr(X \leq 90) \\
&= 1 - \Pr(Z \leq 1.43) \\
&= 1 - 0.92 \\
&= 0.08.
\end{aligned}$$

Thus, we find that 8% of exam scores are larger than 90.

To find the percentage of final exam scores less than 65, we must exploit the symmetry of the normal distribution. Recall that our normal random variable X has mean 70. We are given information about the probability of a standard normal random variable assuming a value less than 0.36, which we have already seen corresponds to the probability of our normal random variable X assuming a value less than 75. Now notice that the x value 65 is the reflection of 75 through the mean of X. That is, both scores 65 and 75 are exactly 5 units from the mean of our random variable X. Thus we can take advantage of the symmetry property of the normal distribution.

Using the symmetry identity from earlier in this section, we find that

$$\begin{aligned}
\Pr(X < 65) &= \Pr(X < 70 - 5) \\
&= \Pr(X > 70 + 5) \\
&= 1 - \Pr(X \leq 75) \\
&= 1 - \Pr(Z \leq 0.36) \\
&= 1 - 0.64 \\
&= 0.36.
\end{aligned}$$

Thus, we find that 36% of exam scores are smaller than 65.

2.5 Expected Value, Variance, and Standard Deviation

Analogous to the discrete case, we can define the expected value, variance, and standard deviation of a continuous random variable. These quantities have the same interpretation as in the discrete setting. The expectation of a random variable is a measure of the center of the distribution, its mean value. The variance and standard deviation are measures of the horizontal spread or dispersion of the random variable.

Expected Value of a Continuous Random Variable

The **expected value** (also called the **expectation** or **mean**) of a continuous random variable X, with probability density function $f(x)$, is the number given by

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

Variance and Standard Deviation of a Continuous Random Variable

The **variance** of a continuous random variable X, with probability density function $f(x)$ is:

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} \left(x - \mathbb{E}(X)\right)^2 f(x) dx.$$

As in the discrete case, the **standard deviation**, σ, is the positive square root of the variance:

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

Simple Example

A random variable X is given by the following PDF. Check that this is a valid PDF and calculate the standard deviation of X.

$$f(x) = \begin{cases} 2(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

Solution

Part 1

To verify that $f(x)$ is a valid PDF, we must check that it is everywhere nonnegative and that it integrates to 1.

We see that $2(1-x) = 2 - 2x \ge 0$ precisely when $x \le 1$; thus $f(x)$ is everywhere nonnegative.

To check that $f(x)$ has unit area under its graph, we calculate

$$\int_{-\infty}^{\infty} f(x)dx = 2\int_{0}^{1}(1-x)dx = 2\left(x - \frac{x^2}{2}\right)\Big|_{0}^{1} = 1$$

So $f(x)$ is indeed a valid PDF.

Part 2

To calculate the standard deviation of X, we must first find its variance. Calculating the variance of X requires its expected value:

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

$$= \int_0^1 x\Big[2(1-x)\Big]\,dx$$

$$= 2\int_0^1 \left(x - x^2\right)dx$$

$$= 2\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1$$

$$= 1/3$$

Using this value, we compute the variance of X as follows

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty}\left(x - \mathbb{E}(X)\right)^2 f(x)\,dx$$

$$= \int_0^1 \left(x - 1/3\right)^2 \cdot 2(1-x)\,dx$$

$$= 2\int_0^1 \left(x^2 - \frac{2}{3}x + \frac{1}{9}\right)(1-x)\,dx$$

$$= 2\int_0^1 \left(-x^3 + \frac{5}{3}x^2 - \frac{7}{9}x + \frac{1}{9}\right)dx$$

$$= 2\left(-\frac{1}{4}x^4 + \frac{5}{9}x^3 - \frac{7}{18}x^2 + \frac{1}{9}x\right)\Big|_0^1$$

$$= 2\left(-\frac{1}{4} + \frac{5}{9} - \frac{7}{18} + \frac{1}{9}\right)$$

$$= \frac{1}{18}$$

Therefore, the standard deviation of X is

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

$$= \frac{1}{3\sqrt{2}}$$

An Alternative Formula for Variance

There is an alternative formula for the variance of a random variable that is less tedious than the above definition.

Alternate Formula for the Variance of a Continuous Random Variable
The **variance** of a continuous random variable X with PDF $f(x)$ is the number given by $$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

The derivation of this formula is a simple manipulation and has been relegated to the exercises. We should note that a completely analogous formula holds for the variance of a discrete random variable, with the integral signs replaced by sums.

Simple Example Revisited

We can use this alternate formula for variance to find the standard deviation of the random variable X defined above.

Remembering that the expectation of X was found to be 1/3, we compute the variance of X as follows:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\, dx - \left(\frac{1}{3}\right)^2$$

$$= 2\int_0^1 x^2(1-x)\, dx - \frac{1}{9}$$

$$= 2\int_0^1 (x^2 - x^3)\, dx - \frac{1}{9}$$

$$= 2\left(\frac{1}{3}x^3 - \frac{1}{4}x^4\right)\Big|_0^1 - \frac{1}{9}$$

$$= 2\left(\frac{1}{3} - \frac{1}{4}\right) - \frac{1}{9}$$

$$= \frac{1}{18}$$

In the exercises, you will compute the expectations, variances and standard deviations of many of the random variables we have introduced in this chapter, as well as those of many new ones.

2.6 A Sample Problem

The length of time X, needed by students in a particular course to complete a 1 hour exam is a random variable with PDF given by

$$f(x) = \begin{cases} k(x^2 + x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{elsewhere} \end{cases}$$

For the random variable X,

1. Find the value k that makes $f(x)$ a probability density function (PDF)
2. Find the cumulative distribution function (CDF)
3. Graph the PDF and the CDF
4. Find the probability that that a randomly selected student will finish the exam in less than half an hour
5. Find the mean time needed to complete a 1 hour exam
6. Find the variance and standard deviation of X

Solution

Part 1

The given PDF must integrate to 1. Thus, we calculate

$$1 = \int_{-\infty}^{\infty} f(x)\,dx$$

$$= k \int_{0}^{1} (x^2 + x)\,dx$$

$$= k \left(\frac{x^3}{3} + \frac{x^2}{2} \right) \Big|_{0}^{1}$$

$$= k \left(\frac{5}{6} \right)$$

Therefore, $k = 6/5$. Notice also that the PDF is nonnegative everywhere.

Part 2

The CDF, $F(x)$, is the area function of the PDF, obtained by integrating the PDF from negative infinity to an arbitrary value x.

If x is in the interval $(-\infty, 0)$, then

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \int_{-\infty}^{x} 0\,dt$$

$$= 0$$

If x is in the interval $[0, 1]$, then

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{x} f(t)\,dt$$

$$= 0 + \frac{6}{5}\left(\frac{x^3}{3} + \frac{x^2}{2} \right)$$

$$= \frac{6}{5}\left(\frac{x^3}{3} + \frac{x^2}{2} \right)$$
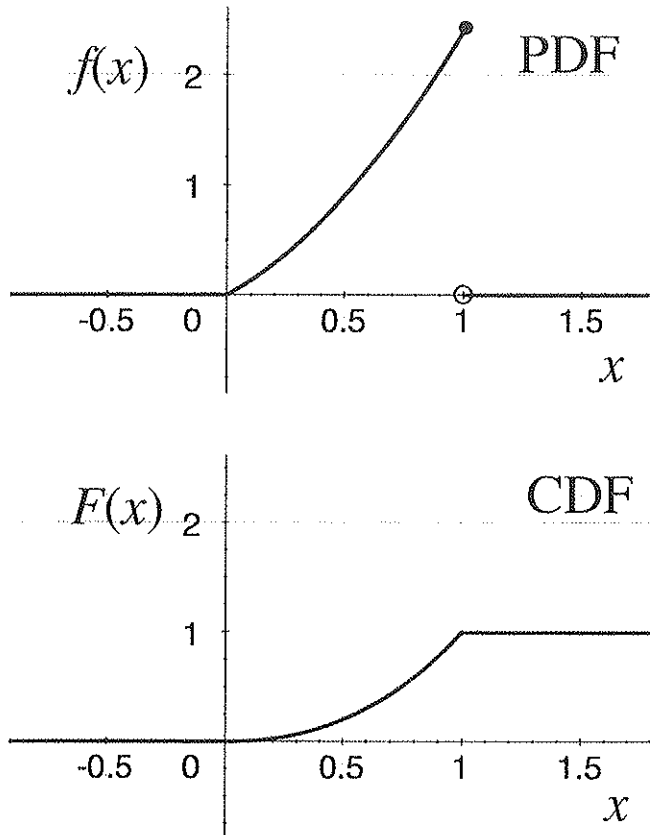
If x is in the interval $(1, \infty)$ then

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{1} f(t)\,dt + \int_{1}^{x} f(t)\,dt$$

$$= 0 + \frac{6}{5}\left(\frac{x^3}{3} + \frac{x^2}{2} \right) \Big|_{0}^{1} + 0$$

$$= \frac{6}{5} \cdot \frac{5}{6}$$

$$= 1$$

The CDF is therefore given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{6}{5}\left(\frac{x^3}{3} + \frac{x^2}{2}\right) & \text{if } 0 \le x \le 1, \\ 1 & \text{if } x > 1. \end{cases}$$

Part 3

The PDF and CDF of X are shown below.

Part 4

The probability that a student will complete the exam in less than half an hour is $\Pr(X < 0.5)$. Note that since $\Pr(X = 0.5) = 0$ (since X is a continuous random variable) it is equivalent to calculate $\Pr(x \le 0.5)$. This is precisely $F(0.5)$:

$$F(0.5) = \frac{6}{5}\left(\frac{0.5^3}{3} + \frac{0.5^2}{2}\right) = \frac{6}{5}\left(\frac{1}{24} + \frac{1}{8}\right) = \frac{6}{5} \cdot \frac{1}{6} = \frac{1}{5}$$

Part 5

The mean time to complete a 1 hour exam is the expected value of the random variable X. Consequently, we calculate

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

$$= \frac{6}{5} \int_0^1 x(x^2 + x) dx$$

$$= \frac{6}{5} \int_0^1 x^3 + x^2 dx$$

$$= \frac{6}{5} \left(\frac{x^4}{4} + \frac{x^3}{3} \right) \Big|_0^1$$

$$= \frac{6}{5} \left(\frac{1}{4} + \frac{1}{3} \right)$$

$$= \frac{7}{10}$$

Part 6

To find the variance of X, we use our alternate formula to calculate

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\frac{7}{10} \right)^2$$

$$= \frac{6}{5} \int_0^1 x^2(x^2 + x) dx - \frac{49}{100}$$

$$= \frac{6}{5} \int_0^1 x^4 + x^3 dx - \frac{49}{100}$$

$$= \frac{6}{5} \left(\frac{x^5}{5} + \frac{x^4}{4} \right) \Big|_0^1 - \frac{49}{100}$$

$$= \frac{6}{5} \left(\frac{1}{5} + \frac{1}{4} \right) - \frac{49}{100}$$

$$= \frac{54}{100} - \frac{49}{100}$$

$$= \frac{1}{20}$$

Finally, we see that the standard deviation of X is

$$\sigma(X) = \sqrt{\frac{1}{20}} = \frac{1}{2\sqrt{5}}$$

2.7 Chapter 2 Summary

Chapter 2 defined continuous random variables and investigated some of their properties. We saw several examples of commonly used continuous distributions, including the famous normal distribution.

Relationship Between CDF and PDF

One of the key features of a random variable is its associated probability distribution, which gives the probabilities that we can observe a certain event, or set of values, under the given random variable. This distribution can take the form of either a cumulative distribution function (CDF) or a probability density function (PDF) for continuous random variables. These two functions are related by the Fundamental Theorem of Calculus:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

The integrand is the PDF of our continuous random variable, and the corresponding integral is the CDF.

Calculating Probabilities

These two functions give the probabilities associated with observing certain events under a random variable X in question. The CDF has a direct probabilistic interpretation, given by

$$F(x) = \Pr(X \leq x)$$

Using the relationship between the CDF and the PDF, probabilities for events associated to continuous random variables can be computed in two equivalent ways. Suppose we wish to calculate the probability that a continuous random variable X is between two values a and b. We could use the PDF and integrate to find this probability.

$$\Pr(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

Alternatively, if we wish to use the CDF, $F(x)$, we can evaluate the difference $F(b) - F(a)$ to find this probability.

$$\Pr(a \leq X \leq b) = F(b) - F(a)$$

Of course we know that both approaches yield the same result. This fact is precisely the statement of the Fundamental Theorem of Calculus.

Expected Value, Variance, and Standard Deviation

Just as with discrete random variables, the expectation represents the "center" of a random variable, an expected value of an experiment, or the average value of the outcomes of an experiment repeated many times. The variance and standard deviation of a random variable is a numerical measure of the spread, or dispersion, of the PDF of X. Given the PDF $f(x)$ of a continuous random variable X, we can calculate these quantities.

We collect the formulas for the expected value, variance, and standard deviation of a continuous random variable X with PDF $f(x)$ in the following table.

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

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x)\,dx$$

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$