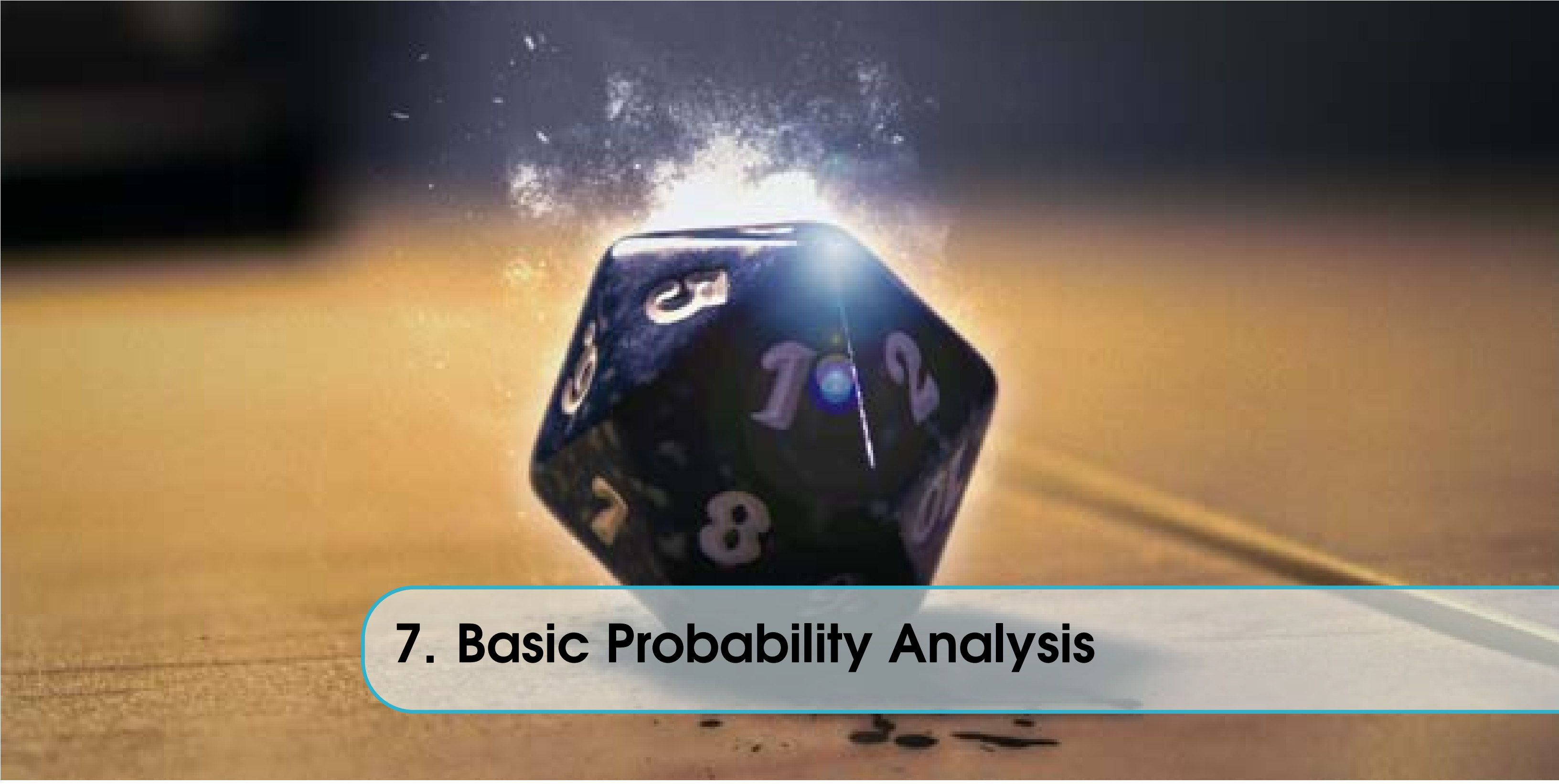

7. Basic Probability Analysis

The study of probability is all about the study of chance – the quantification of the chance or the likelihood that something (e.g. an event, an outcome) will happen. In this chapter we will recall some elementary concepts in probability that are very much required to perform system modeling and analysis.

7.1 Events and their Probabilities

Probability is usually measured in fractions or percentages. For example, a weather forecaster might say that there is a one in five chance (or probability) of rain. This probability is expressed as one fifth or 20%. As another example, consider the outcome of rolling a die. The die has six faces, numbered as 1, 2, 3, 4, 5, and 6. Since it is equally likely that the die will end up with any one of these faces up, the likelihood (or probability) that the outcome is any one of these values is precisely 1/6.

In both of the above examples (weather forecast and rolling a die), we associated a probability with a specific event which is the outcome of a specific experiment. For example, the experiment at hand may be checking whether it is raining at a particular locale. The outcome of this experiment is either "raining" or "not raining." Similarly, the experiment may be rolling a die, in which case the outcome of the experiment is a number between 1 and 6.

The maximum probability of any outcome is 1 or 100%. A probability of 1 represents certainty, meaning that the outcome will definitely happen. For example, the probability that the sun will set within the next 24-hour period (in Boston) is 1. A probability of 0 represents certainty as well! It means that the outcome associated with that 0 probability will never happen. The probability of obtaining 7 as the outcome of rolling a six-sided die is 0.

To be able to evaluate the likelihood of an event, one needs to figure out the number of ways that the event will materialize relative to all possible outcomes of the experiment. Example: A disk cache holds 50 blocks of data. Assuming that the disk contains 10,000 such blocks, what is the probability that a request to the disk will hit in the disk cache? Solve this problem under the assumption that all disk blocks are equally likely to be accessed.

7.1.1 A Motivating Example

In a class of N students, what is the probability that two students will have the same birthday?

The above question is one of the "classical" questions used to motivate the concepts of event probabilities. To answer this question, we must first make some assumptions so that we are able to give an exact answer[1]. Specifically, we must assume that the likelihood of birth on any day of the year is the same and that any year consists of exactly 365 days (no leap years).

Now we are ready to answer the question.

Case 1: Consider the case when $N = 2$ (i.e. two people). First, we must ask ourselves how many ways the two birthdays can occur throughout the year. The answer is that there are 365 ways for the first birthday and for each of these there are 365 ways for the second birthday. The total is thus $365 \cdot 365 = 365^2$.

Now we ask another question: How many ways could the two birthdays occur so that they are not on the same day? The answer is to let the first birthday be on any day (a total of 365 possibilities) and to restrict the second birthday to be on one of the remaining 364 days. Thus the number of ways the two birthdays could occur on different days is $365 \cdot 364$. This leads to the probability of the two birthdays being on different days being $\frac{364 \cdot 365}{365^2} = 0.997$. Thus, the probability that the two birthdays will coincide is: $1 - \frac{364 \cdot 365}{365^2} = 0.003$.

Case 2: Consider the case when $N = 3$ (i.e. three people). First, we must ask ourselves how many ways the three birthdays can occur throughout the year. The answer is that there are 365 ways for the first, 365 ways for the second and 365 for the third. The total is thus $365 \cdot 365 \cdot 365 = 365^3$.

Now we ask another question: How many ways could the three birthdays occur so that they are not on the same day? The answer is to let the first birthday be on any day (a total of 365 possibilities) and to restrict the second birthday to be on one of the remaining 364 days, and the third birthday to be on one of the remaining 363 days. Thus, the number of ways the three birthdays could occur on different days is $365 \cdot 364 \cdot 363$. This leads to the probability of the two birthdays being on different days being $\frac{363 \cdot 364 \cdot 365}{365^3} = 0.992$. Thus, the probability that the two birthdays will coincide is: $1 - \frac{363 \cdot 364 \cdot 365}{365^3} = 0.008$.

Now we are ready to generalize this to N.There are 365^N ways of choosing N birthdays. Of these, there are only $\frac{365!}{(365-N)!}$ ways of making the N birthdays fall on distinct days. Thus, the probability of

[1]Making assumptions is a hallmark of trying to compute the probabilities of events in the real world. As a computer scientist, you should always be willing to make assumptions (and explain why they are good ones to make), and you should always question assumptions that other people may make! These assumptions are typically written in fine print like this footnote!

(at least) two birthdays being on the same day is:

$$1-\frac{\frac{365!}{(365-N)!}}{365^N}=1-\frac{365\cdot 364\cdot\ldots\cdot(365-N+1)}{365^N} \tag{7.1}$$

The plot in Figure 7.1 below shows the value of this probability for different values of N.

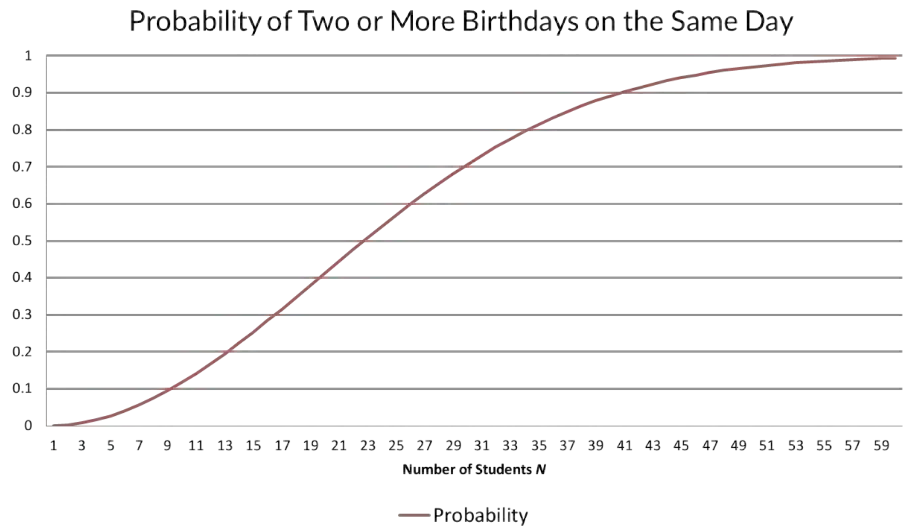

Figure 7.1: Probability of two birthdays falling on the same calendar day in a year as function of the number of considered students N.

Example: A web server has 500 different files. Assuming that all files are equally likely to be accessed (i.e. all files are equally popular), what is the probability that at least 2 out of 10 hits to the web server will be for the same file? What is that probability that at least 2 out of 100 hits to the web server will be for the same file? What is the general solution for the probability that at least 2 out of N hits to the web server will be for the same file? Please note the strong similarities of this problem with what we just solved above talking about students and birthdays.

7.2 Dealing with Multiple Events

Events we described so far are quite simple; they are associated with the outcome of a single experiment. One can think of much more complicated experiments! For example, one can think of an experiment in which we roll a pair of dice and report the sum of the two dice. Or, one can think of two experiments, whereby in the first we measure the CPU utilization and in the second we measure the throughput of the system.

7.2.1 Conditional Probability

Consider two events A and B. For example, event A may be the event that there are more than 100 processes running in the system and event B may be the event that the system hangs. For each one of these events, we can compute (or we may be given) the probability of occurrence. Let $P(A)$ and $P(B)$ be these probabilities. Thus $P(A)$ is the probability that event A occurs and $P(B)$ is the probability that event B occurs. In addition to $P(A)$ and $P(B)$, we may also be given the probability that **both** events will occur (the probability of the joint event). Let $P(A \cap B)$ denote that probability.

Now consider the following question: "If we know that event A did occur, what is the probability that event B will occur?" The more "formal" way of posing this question is: "What is the conditional probability of B given A?" We denote this by $P(B|A)$.

So, using our example of A and B, the question is "What is the conditional probability that the system will hang given that the system has more than 100 processes?"

Notice that this question is different from "What is the probability that the system hangs?" – which is precisely $P(B)$ – because knowing that the system has 100 processes may change our expectation of whether the system will hang or not.

To answer this question, we rely on a very important result in probability – namely **Baye's Rule**. Baye's rule states the following.

Theorem 7.2.1 — Baye's Rule.
The probability that two events A and B jointly occur, expressed as $P(A \cap B)$, can be computed as:

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \tag{7.2}$$

where $P(A)$ is the probability that A occurs, $P(B)$ is the probability that B occurs, $P(A|B)$ is the probability that A occurs given that B occurred, and $P(B|A)$ is the probability that B occurs given that A occurred.

So, for example, if we are told that:

- $P(A)$ = P("Number of processes in the system = 100") = 0.005,
- $P(B)$ = P("System hangs") = 0.01, and
- $P(A \cap B)$ = P("Number of processes in the system=100 and the system hangs") = 0.001,

then we can compute the probability that the system will hang (B) provided that it has 100 processes (A). This corresponds to $P(B|A) = P(A \cap B)/P(A) = 0.001/0.005 = 0.2$.

Notice that Baye's rule allow us also to turn a conditional probability of event B given event A into the conditional probability of event A given event B.

So, for example, if we are told that:

- $P(A)$ = P("Raining in Boston") = 0.15,
- $P(B)$ = P("Red Sox winning a game") = 0.6, and
- $P(B|A)$ = P("Red Sox winning given that it is raining in Boston") = 0.8,

then we can compute the probability that it will rain given that the Red Sox won a game i.e.

$P(A|B) = P(B|A) \cdot P(A)/P(B) = 0.8 \cdot 0.15/0.6 = 0.2.$

Example: The probability that an access to memory will hit in the cache is 0.95. The probability that an access to memory will hit in the cache given that the previous access hit in the cache is 0.99. What is the probability that two accesses in a row will hit in the cache? What is the probability that an access to memory was a hit given that the next access was a hit?

7.2.2 Independence

Another important aspect of Baye's rule is that it allows us to think about a very important concept—namely that of independence. Two events are deemed independent if the occurrence of the first event bears no significance on the probability of occurrence of the second event. This implies the following fact (directly evident from the definition of conditional probability).

Theorem 7.2.2 — Event Independence.
If A and B are independent events, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

For example, if there is no relationship between two events (say two successive coin tosses) then the probability of getting heads on the second coin toss given that we got (say) tails on the first coin toss will be still 0.5. Thus P(2nd toss is tails | 1st toss is heads) = P(tails) = 0.5. For independent events A and B, we have $P(A|B) = P(A)$ and $P(B|A) = P(B)$ (see Theorem 7.2.2). Substituting in Baye's rule, we obtain the following important result for independent events.

Corollary 7.2.3 — Conditional Probability of Independent Event.
If A and B are independent events, then $P(A \cap B) = P(A) \cdot P(B)$.

7.3 Union and Intersection of Events

As we hinted before, we may be interested in complex events that involve the outcome of multiple experiments or observations. In order to evaluate the probability of a specific outcome of a "complex" experiment, one has to find out the many ways such an outcome could be produced and then evaluate the probability of each such outcome. For example, we might want to evaluate the probability that the outcome of rolling a pair of dice is "4," as highlighted in Figure 7.2. We will find that we can get this outcome if the first die is a "1" and the second die is a "3" (as highlighted below), or if the first die is a "2" and the second die is a "2," or if the first die is a "3" and the second die is a "1."

In general, a complex event can be "decomposed" into simpler events that are either conjoined (i.e. must happen together) or disjoint (i.e. exclusive).

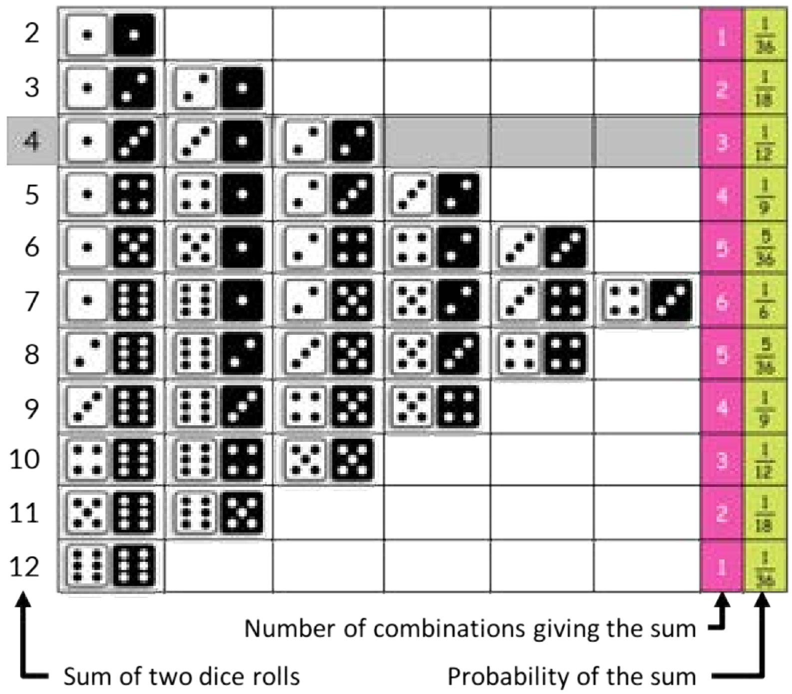

Figure 7.2: Complete table of all the possible outcomes and their probabilities for the sum of a two-dice roll.

7.3.1 Joint Events

The probability of getting a “1” on the first die is 1/6 and the probability of getting a “3” on the second die is also 1/6. Since the outcome of the second die was independent of the outcome of the first (i.e. the outcome of one die did not influence the outcome of the second), we can conclude that the probability of the *joint event* (“1” on first die and “3” on second die) is $(1/6) \cdot (1/6) = 1/36$.

It is extremely important to notice that we are able to compute the joint event probability by multiplying the probabilities of the two events only because the two events were **independent**[2] , as we explained before.

7.3.2 Mutually Exclusive (Disjoint) Events

Now, let’s go back to calculating the probability of getting a total of “4” when rolling the pair of dice. Recall that we figured out that the probability of a “1”on the first die and a “3” on the second die is 1/36. Notice that getting a “1” on the first die and a “3” on the second die is only one of the several different ways of getting an outcome of “4.” We can also get “4” if we get a 2 on the first die and a 2 on the second die. In addition, we can get a “4” if we get a “3” on the first die and a “1” on the second die.

[2] If the two events were not independent, we cannot simply “multiply” their probabilities! Rather, we would have had to consider their conditional probabilities as stated by the Baye’s Rule.

Each one of these possibilities (1+3, 2+2, or 3+1) is exclusive of the other possibilities. Since the outcome we desire is obtained if any one of these disjoint events occurs, we calculate the probability to be the sum of these three possibilities = 1/36 + 1/36 + 1/36 = 1/12.

In general, if we are interested in the probability that one of two events A or B, expressed as $P(A \cup B)$ will occur, we have to use the following relationship:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{7.3}$$

and the reason we can simply add the probabilities of mutually exclusive events is because for such events $P(A \cap B) = 0$.

7.3.3 Complementary Events

Two mutually exclusive events are complementary if the probability of their union is 1. Thus, if A and B are two events satisfying $P(A \cap B) = 0$, and $P(A) + P(B) = 1$, then they are complementary. We typically refer to the complement of an event A by $\bar{A}$.

Notice that mutually exclusivity (and by association complementary events) implies dependence! In fact, one can show that for two mutually exclusive events A and B, $P(A|B) = 0$. Also, one can show that for two mutually exclusive events A and B, $P(A|\bar{B}) >= P(A)$.

7.4 Random Variables

In the previous example (of rolling a pair of dice), the outcome of the experiment (the sum of the two dice) could be a number ranging from 2 to 12. This is an example of a "random variable" because the value of that variable is not known until the experiment is conducted (e.g. we roll the dice).

A **random variable** is a variable whose value could be one of a (possibly infinite) number of values. Examples of random variables include:

- The outcome of a coin toss
- The outcome of rolling a die
- The outcome of rolling a pair of dice
- The height or the weight of a student at Boston University
- The number of students registered in CS-350 in a particular semester
- The grade of a student in given class
- The number of packets that are transmitted on a LAN in a given period of time
- The number of blocked processes at given point in time
- The amount of time a process waits for an event
- The interval of time between two failures
- The interval of time it takes to recover from a failure
- The number of collisions detected on an Ethernet LAN per unit time
- The time it takes to access a hierarchical memory

In each one of the above cases, we don't know a priori what would be the value of such a variable (height, weight, students registered, grade, number of packets, number of blocked processes, waiting time for an event, etc.) We may have a pretty good idea about the "likelihood" of a particular value or range of values, but we cannot foretell what a specific instance we would observe would be.

Another way of thinking of this is to view a random variable as a reflection of some quantity in the real world that can take on a number of different values with different "probabilities." More technically, a random variable is a mapping from the set of all possible events under consideration to real numbers. That is, a random variable associates a real number with each event. This concept is sometimes expressed in terms of an experiment with many possible outcomes; a random variable assigns a value to each such outcome.

Random variables may be **continuous** or **discrete**. A random variable is continuous if it takes on a non-countable infinite number of distinct values. A random variable is discrete if it takes on a finite or countable infinite number of values. For example, height, weight, and waiting time are continuous, whereas the number of students, number of packets, and grades are discrete.

7.4.1 Continuous and Discrete Random Variables

A continuous random variable could assume any one of an infinite, uncountable number of real values. Examples of continuous random variables include the height of an individual or the time interval between two keystrokes on a keyboard.

A discrete random variable could assume any one of possibly infinite, but countable number of integer values. Some example of a discrete random variable include the number of people in an elevator, the number of processes in the ready queue of an OS, or the number of people crossing a bridge in any 1-minute interval.

7.4.2 Probability

While one could not know *a priori* the specific value that a random variable will assume, one could calculate (or measure) the likelihood of various values. For example, while one could not predict a priori the height of a student selected at random from the Boston University population, one could assume that the height of such student is unlikely to be less than (say) 5 feet or more than (say) 6 feet. Moreover, it is very unlikely for such random variable to be less than 4 feet (or more than 7 feet).

For discrete random variables, the probability that a random variable will assume a particular value is a measure of how "likely" it is for that random variable to assume that value.

For continuous random variables, we cannot speak of the probability of the random variable assuming a specific "exact" value (because the probability of that happening is zero). Rather, we consider the likelihood of the random variable being within a range of values. For example, when we talk about heights, we would talk about the probability of a student being between 5'9" and 5'10" tall.

One way of capturing the concept of "probability" is to imagine that we have the opportunity to observe the random variable for an *infinite* number of times. The probability that the random variable will assume a particular value (or will lie in a given range of values) is equal to the ratio between the number of times that value (or range) was observed and the total number of observations.

The word "infinite" is a keyword in the previous paragraph. In particular, if we observe the random variable only a finite number of times, then our evaluation of the probability of a specific outcome is likely to be biased! For example, if we select 10 students at random from BU and measure their heights, we may manage to get a biased "sample" of 6-foot-and-taller students, which may bias our estimation of the probability of a BU student being of a certain height.

So, how do we observe a random variable an infinite number of times? This is obviously impractical. There are two ways to deal with this.

First, one can avoid having to observe the random variable "forever" before making conclusions about probability of outcomes by analyzing the "experiment" or by inspecting the entire population (if the population is finite). An example of this is characterizing the probability of the outcome of a dice roll experiment we discussed earlier.

Second, (if neither analysis nor inspection of the whole population is feasible), the only hope is to **approximate** what the probabilities of specific outcomes will be based on a sufficiently large number of observations (or samples). An example of this is characterizing the probability of a BU student being of a specific height or weight.

The second of the above approaches is what statistics is all about – answering questions about a population using finite observations or measurements.

7.5 Probability Distributions

The **probability distribution** of a random variable is the characterization of "the probability of that random variable assuming all of its possible outcomes." In other words, if we know the probability that a random variable will assume a particular value (or will lie within a continuum of values between an upper and lower bound), then we have characterized the random variable's probability distribution.

Going back to our earlier example of rolling a pair of dice and observing the result, we can characterize the random variable S denoting the sum of the two dice by listing the probability of every possible outcome. If we do so, we get the probability for every possible outcome of $S = 2, 3, 4, \ldots, 12$. This is shown in the graph below, Figure 7.3, and goes under the name of **Probability Mass Function** for S.

The above plot shows the probability that S assumes a specific value. We can also think of the probability that S will assume a value that is less than a specific value x. This is simply the sum of the probabilities for $S = 2, 3, \ldots, x$. This is shown in the graph below, Figure 7.4, called the **Cumulative Distribution Function** or **Probability Distribution Function** for S.

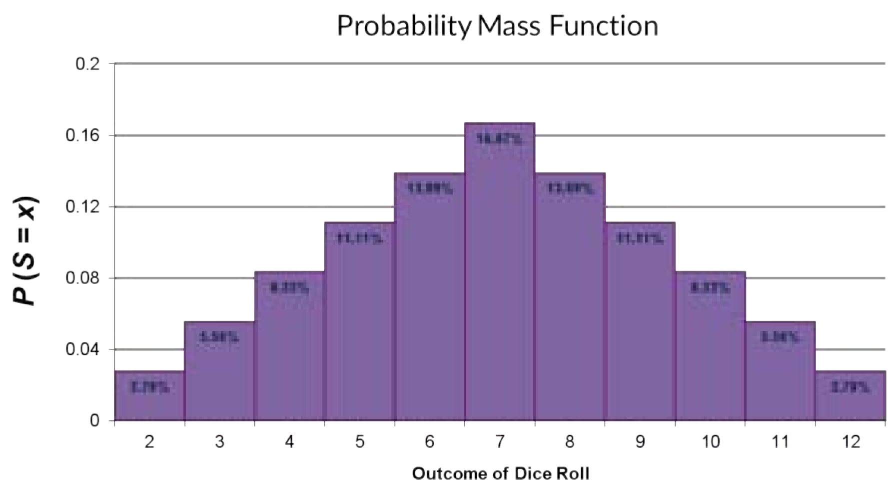

Figure 7.3: Probability Mass Function of the random variable S used to capture the result of a two-dice roll.

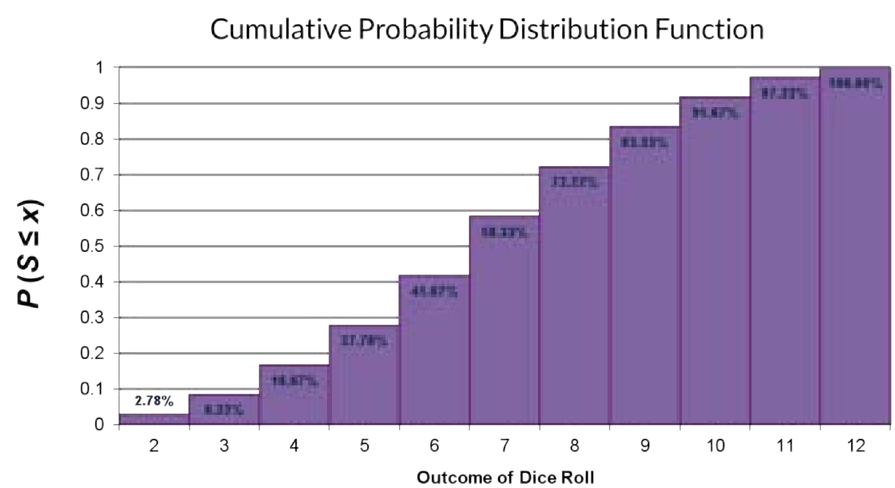

Figure 7.4: Cumulative Distribution Function of the random variable S used to capture the result of a two-dice roll.

We are now ready to define the probability mass and distribution functions for a random variable.

7.5.1 Probability Mass (Density) Function

One way of characterizing a random variable is to provide a function that gives the probability that the random variable will be *equal to a particular value* (for discrete random variables) or will *lie within an infinitely small range of values* (for continuous random variables).

For a discrete random variable X, we use the notation $f(x)$ to describe the *Probability Mass Function* (PMF) of X. Namely, $f(x)$ is the probability that the random variable X is equal to x. In other words:

$$f(x) = P(X = x) \tag{7.4}$$

For a continuous random variable X, for which point values have zero probabilities, rather than using a probability mass function, we use a *Probability Density Function* (PDF) $f(x)$, which when integrated over an interval yields the probability that the continuous random variable will be in that interval.

7.5.2 Cumulative (Probability) Distribution Function

Another way of characterizing a random variable (whether discrete or continuous) is to provide a function that gives the probability that the random variable will be *less than a particular value*. For a random variable X, we use the notation $F(x)$ to describe the *Cumulative Distribution Function* (CDF) of X. $F(x)$ is the probability that the random variable X is less than or equal to x. The cumulative distribution function of X is also often called the Probability Distribution Function of X. In other words:

$$F(x) = P(X \leq x) \tag{7.5}$$

Obviously, $F(-\infty) = 0$. This means that we are "certain" that the random variable cannot be smaller than $-\infty$. Similarly, $F(+\infty) = 1$, meaning that we are "certain" that the random variable is smaller than infinity. A qualitative depiction of a probability mass (density) function $f(x)$ (left-hand side) and of a cumulative distribution function $F(x)$(right-hand side) is provided in Figure 7.5.

7.5.3 Relationship Between $f(x)$ and $F(x)$

For discrete random variables, we can relate $F(x)$ and $f(x)$ as:

$$F(x) = \sum_{y=-\infty}^{x} f(y) \tag{7.6}$$

For continuous random variables, instead, it holds that:

$$F(x) = \int_{-\infty}^{x} f(y)dy \tag{7.7}$$

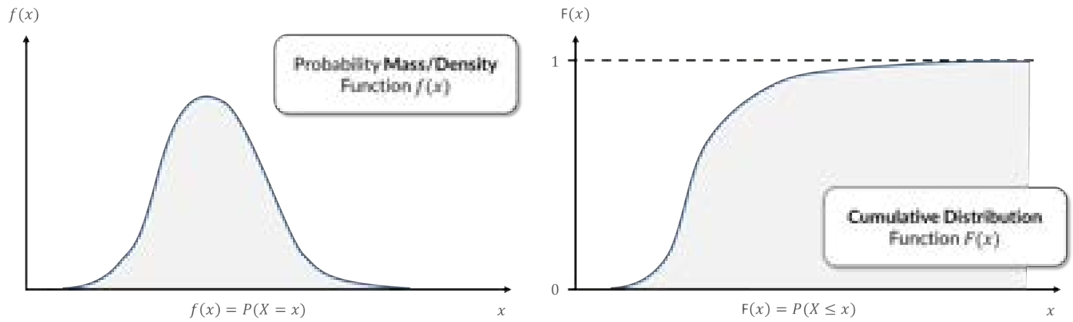

Figure 7.5: Probability mass (density) function $f(x)$ (left-hand side) and cumulative distribution function $F(x)$(right-hand side) for a generic random variable X.

The cumulative probability distribution function could be used to compute the probability that a random variable lies within a range of values. In particular, for discrete variables, $P(x_1 < X \leq x_2)$ is given by:

$$F(x_2) - F(x_1) = \sum_{y=-\infty}^{x_2} f(y) - \sum_{y=-\infty}^{x_1} f(y) = \sum_{y=x_1}^{x_2} f(y) \tag{7.8}$$

And similarly, for continuous variables, we have:

$$F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f(y)dy - \int_{-\infty}^{x_1} f(y)dy = \int_{x_1}^{x_2} f(y)dy \tag{7.9}$$

In the discussion so far, we have been careful to distinguish between the "random variable" and the "value that the random variable can assume." Namely, we have used upper-case letters to denote random variables and lower-case letters to denote values of these random variables. For convenience, in the remainder of this document, we will no longer make that distinction. However, it should be noted that a distinction exists (which we assume will be clear from the context).

Also, unless otherwise noted, in the remainder of this text, whenever we describe a random variable, we will assume it is continuous[3].

7.5.4 Random Expressions

An expression that contains one or more random variables is itself a random variable. Thus, if x and y are random variables, then $2xy$, x^2, $\log(x)$, ye^x, etc. are all random variables that are expressed as a function of the random variables x and/or y.

The following are examples of random expressions:

- The difference between the heights of two BU students (the difference between two independent and identically distributed (i.i.d.) random variables for the height of BU students);

[3]The case for the discrete random variable could be easily obtained by replacing every integration by a summation.

- The time from when a process is added to the ready queue until it finishes using the CPU (the total of the random variables for waiting in the queue and service time on the CPU);
- The availability of a computing system (a function of the random variables describing the interarrival of failures and the repair times).

7.6 Expected Value of a Random Variable

Let us first define the expected value of a random variable x.

Definition 7.6.1 — Expected Value.
The expected value of any function $g(x)$ of a random variable x with probability density $f(x)$ is defined by:

$$E[g(x)] = \int_{-\infty}^{+\infty} g(y) \cdot f(y) dy \tag{7.10}$$

Using the above relationship, one can easily show the following important properties:

Observation 7.6.1 — Properties of Expected Value.
The following properties hold:

1. If a is a constant then $E[a \cdot g(x)] = a \cdot E[g(x)]$;
2. $E[g_1(x) + g_2(x)] = E[g_1(x)] + E[g_2(x)]$;
3. $E[g(x) + h(y)] = E[g(x)] + E[h(y)]$;
4. If x and y are **independent** random variables then $E[g(x) \cdot h(y)] = E[g(x)] \cdot E[h(y)]$.

The first relationship (1) says that the expected value of (say) twice a random variable is the same as twice the expected value of the random variable. The second and third relationships (2 and 3) imply that the expected value of the sum of two functions is the same as the sum of the expected value of each one of the two functions alone. The fourth relationship (4) says that the expected value of the product of two functions (in different random variables) is equal to the product of the expected value of these functions, so long as the random variables are independent.

7.6.1 Moments of a Random Variable

Of particular interest in describing a random variable x are the expected values of the expressions x, x^2, x^3, etc. These are expressed as $E(x), E(x^2), E(x^3)$, etc., which we call the **moments** of the random variable x.

7.6.2 Mean, Variance and Standard Deviation

The **mean** μ of a random variable x is defined as the expected value of that random variable $E(x)$, i.e. its first moment. It is given by:

$$\mu = E[x] = \int_{-\infty}^{+\infty} y \cdot f(y) dy \tag{7.11}$$

Often times we are interested in the *variability* of the random variable and thus we are interested in computing the expected value of the difference between the random variable and its mean without consideration to the sign of the difference. To do so, we consider the squared difference between a random variable and its mean.

The **variance** var$[x]$ of a random variable x is defined as the expected value of the squared difference between a random variable and its mean. The **standard deviation** σ is the square root of the variance.

$$\sigma = \sqrt{\text{var}[x]} = \sqrt{E[(x-\mu)^2]} \tag{7.12}$$

One can easily show that the standard deviation can also be calculated using the following equation:

$$\sigma = \sqrt{E[x^2] - \mu^2} \tag{7.13}$$

To do so, notice that $E[(x-\mu)^2] = E[x^2 + \mu^2 - 2x\mu] = E[x^2] + \mu^2 - 2\mu E[x]$. But since $E[x] = \mu$ we have $E[(x-\mu)^2] = E[x^2] - \mu^2$, and Equation 7.13 follows.

7.6.3 Median and Quantiles

Another measure of interest is the value of the random variable for which the probability distribution function is 0.5. This value is called the **median** because the probability that a random variable selected from that distribution will be less than the median is exactly 0.5. Thus, the median is the value of the random variable for which half of the observations are smaller and the other half is larger. For example, if the median of the random variable representing the response time for accessing a web page is 1.5 seconds, then we expect that 50% of the accesses to the web page will be completed in less than 1.5 seconds and 50% of the accesses will take longer than 1.5 seconds.

If the median is smaller than the mean, then the probability distribution is said to have a *positive skew*. A distribution that has a positive skew is one that characterizes a random variable for which small values are highly probable, but very large values could occur. An example is depicted in Figure 7.6 on the left-hand side. An example of such distributions is the distribution that characterizes file sizes for WWW documents. The (very) positive skew of Web file size distributions (see work we have done within the Oceans research group in the department) suggests that if we represent the size of web documents as a random variable, then while there are many very large documents (say > 10 MB in size), the probability of small documents (say < 10 KB) is quite large.

If the median is larger than the mean, then the distribution is said to have a *negative skew*. A distribution that has a negative skew is one that characterizes a random variable for which large values are highly probable, but in which very small values could occur. An example is depicted in the right-hand side of Figure 7.6. Typically, an example of this would be the grades of students on a homework assignment (most people do very well, but there are very bad scores as well).

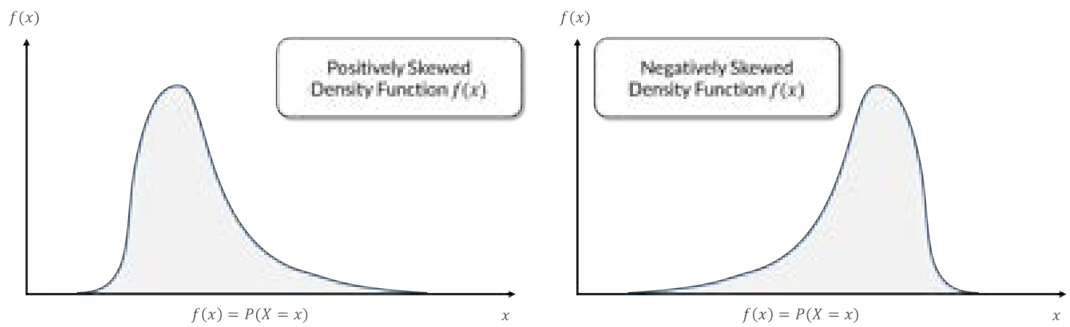

Figure 7.6: Qualitative example of a probability density function $f(x)$ with positive (left-hand side) and negative (right-hand side) skew.

A generalization of the notion of *median* is that of **quantiles**. The $Q\%$-quantile for a random variable x is the value for which $F(x) = Q/100$. Thus, the 99%-quantile is the value of the random variable x for which $F(x) = 0.99$[4].

[4] Obviously the 50% percentile corresponds to the median!