ECE 302: Probability and Applications[1]

Week 3 Topics

- Independence of Events
- Sequential Experiments
  - Independent Bernoulli Trials
  - Binomial Probabilities
  - Geometric Probabilities
- Notion of a Random Variable
  - Cumulative Distribution Function
  - Three types of Random Variables

1 Independence of Events

If the occurrence of event B does not change the probability of event A, then it is natural to say that event A does not depend on B.

- In terms of probabilities this situation occurs when the probability of A given B does not change, that is,

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

- In the above equation we need that $P[B] = 0$.

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

  $$P[A \cap B] = P[A]P[B] \tag{1}$$

- Equation (1) then implies both of the following:

  $$P[A \mid B] = P[A] \tag{2}$$

and

$$P[B \mid A] = P[B] \tag{3}$$

- Note that Eq. (2) implies Eq. (1) when $P[B] \neq 0$ and Eq. (3) implies Eq. (1) when $P[A] \neq 0$.

Example 1

A ball is selected from an urn containing two black balls, numbered 1 and 2, and two white balls, numbered 3 and 4. Let the events A, B, and D be defined as follows:

$$\begin{aligned}
A &= \{(1,b),(2,b)\}, && \text{"black ball selected";} \\
B &= \{(2,b),(4,w)\}, && \text{"even-numbered ball selected"; and} \\
C &= \{(3,w),(4,w)\}, && \text{"number of ball is greater than " 2.}
\end{aligned}$$

Are events A and B independent? Are events A and C independent?

First, consider events A and B. The probabilities required by Eq. (1) are

$$P[A] = P[B] = \frac{1}{2}$$

$$P[A \cap B] = P[\{(2, b)\}] = \frac{1}{4}$$

Thus

$$P[A \cap B] = \frac{1}{4} = P[A]P[B]$$

and the events A and B are independent. Equation (3) gives more insight into the meaning of independence:

$$\begin{aligned} P[A \mid B] &= \frac{P[A \cap B]}{P[B]} = \frac{P[\{(2, b)\}]}{P[\{(2, b), (4, w)\}]} = \frac{1/4}{1/2} = \frac{1}{2} \\ P[A] &= \frac{P[A]}{P[S]} = \frac{P[\{(1, b), (2, b)\}]}{P[\{(1, b), (2, b), (3, w), (4, w)\}]} = \frac{1/2}{1} \end{aligned}$$

These two equations imply that $P[A] = P[A \mid B]$ because the proportion of outcomes in S that lead to the occurrence of A is equal to the proportion of outcomes in B that lead to A. Thus knowledge of the occurrence of B does not alter the probability of the occurrence of A.

Events A and C are not independent since $P[A \cap C] = P[\varnothing] = 0$ so

$$P[A \mid C] = 0 \neq P[A] = .5$$

In fact, A and C are mutually exclusive since $A \cap C = \varnothing$, so the occurrence of C implies that A has definitely not occurred.

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

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

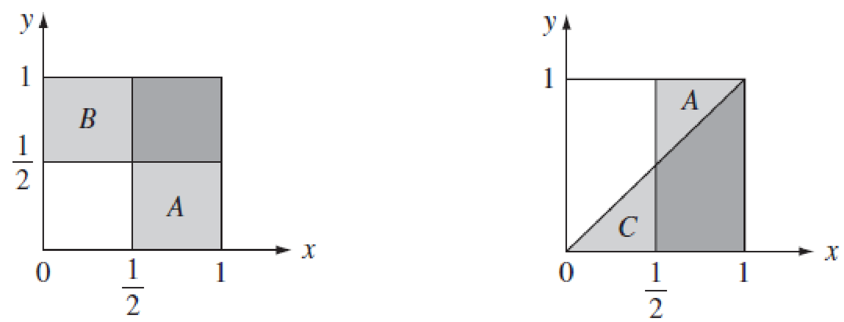
  which implies that at least one of the events must have zero probability.

Example 2

Two numbers x and y are selected at random between zero and one. Let the events A, B, and C be defined as follows:

$$A = \{x > 0.5\}, \quad B = \{y > 0.5\}, \quad \text{and } C = \{x > y\}$$

Are the events A and B independent? Are A and C independent?



(a) Events A and B are independent. **(b)** Events A and C are not independent.

Figure 1: Examples of independent and nonindependent events

Figure 1(a and b) shows the regions of the unit square that correspond to the above events. Using Eq. (3), we have

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} = \frac{1/4}{1/2} = \frac{1}{2} = P[A],$$

so events A and B are independent. Again we have that the "proportion" of outcomes in S leading to A is equal to the "proportion" in B that lead to A.

Using Eq. (2), we have

$$P[A \mid C] = \frac{P[A \cap C]}{P[C]} = \frac{3/8}{1/2} = \frac{3}{4} \neq \frac{1}{2} = P[A]$$

so events A and C are not independent. Indeed from Fig. 1(b) we can see that knowledge of the fact that x is greater than y increases the probability that x is greater than 0.5.

What conditions should three events A, B, and C satisfy in order for them to be independent? First, they should be pairwise independent, that is,

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

In addition, knowledge of the joint occurrence of any two, say A and B should not affect the probability of the third, that is,

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

For this to hold, we must have

$$P[C \mid A \cap B] = \frac{P[A \cap B \cap C]}{P[A \cap B]} = P[C].$$

This in turn implies that we must have

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

where we have used the fact that A and B are pairwise independent. Thus we conclude that three events A, B, and C are independent if the probability of the intersection of any pair or triplet of events is equal to the product of the probabilities of the individual events.

- The following example shows that if three events are pairwise independent, it does not necessarily follow that $P[A \cap B \cap C] = P[A]P[B]P[C]$.

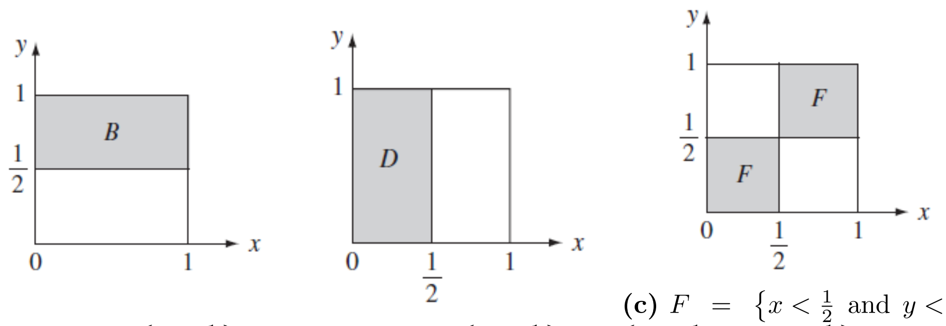
Example 3

Consider the experiment discussed in Example 2 where two numbers are selected at random from the unit interval. Let the events B, D, and F be defined as follows:

$$B = \left\{ y > \frac{1}{2} \right\}, \quad D = \left\{ x < \frac{1}{2} \right\}$$
$$F = \left\{ x < \frac{1}{2} \text{ and } y < \frac{1}{2} \right\} \cup \left\{ x > \frac{1}{2} \text{ and } y > \frac{1}{2} \right\}.$$

The three events are shown in Fig. 2. It can be easily verified that any pair of these events is independent:

$$P[B \cap D] = \frac{1}{4} = P[B]P[D]$$



(a) $B = \{>> \frac{1}{2}\}$ **(b)** $D = \{x < \frac{1}{2}\}$ **(c)** $F = \{x < \frac{1}{2} \text{ and } y < \frac{1}{2}\}$ $\{x > \frac{1}{2} \text{ and } y > \frac{1}{2}\}$

Figure 2: Events B, D, and F are pairwise independent, but the triplet B, D, F, are not independent events.

$$
\begin{aligned}
P[B \cap F] &= \frac{1}{4} = P[B]P[F], \text{ and} \\
P[D \cap F] &= \frac{1}{4} = P[D]P[F]
\end{aligned}
$$

However the three events are not independent, since $B \cap D \cap F = \varnothing$, so

$$
P[B \cap D \cap F] = P[\varnothing] = 0 \neq P[B]P[D]P[F] = \frac{1}{8}
$$

In order for a set of n events to be independent, the probability of an event should be unchanged when we are given the joint occurrence of any subset of the other events. This requirement naturally leads to the following definition of independence. The events $A_1, A_2, \ldots, A_n$ are said to be **independent** if for $k = 2, \ldots, n$,

$$
P\left[A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right] = P\left[A_{i_1}\right] P\left[A_{i_2}\right] \ldots P\left[A_{i_k}\right]
$$

where $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

- For a set of n events we need to verify that all $2^n - n - 1$ possible intersections factor in the right way.

- The above definition of independence appears quite cumbersome because it requires that so many conditions be verified.

- The most common application of the independence concept is in making the assumption that the outcomes of separate experiments are independent. For example it is common to assume that the outcome of a coin toss is independent of the outcomes of all prior and all subsequent coin tosses.

Example 4

Suppose a fair coin is tossed three times and we observe the resulting sequence of heads and tails. Find the probability of the elementary events.

The sample space of this experiment is $S = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH},$ TTH, THT, HTT, TTT $\}$. The assumption that the coin is fair means that the outcomes of a single toss are equiprobable, that is, $P[\text{H}] = P[\text{T}] = 1/2$.

If we assume that the outcomes of the coin tosses are independent, then

$$\begin{aligned}
P[\{\text{HHH}\}] &= P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{H}\}] = \frac{1}{8}, \\
P[\{\text{HHT}\}] &= P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{T}\}] = \frac{1}{8}, \\
P[\{\text{HTH}\}] &= P[\{\text{H}\}]P[\{\text{T}\}]P[\{\text{H}\}] = \frac{1}{8}, \\
P[\{\text{THH}\}] &= P[\{\text{T}\}]P[\{\text{H}\}]P[\{\text{H}\}] = \frac{1}{8}, \\
P[\{\text{TTH}\}] &= P[\{\text{T}\}]P[\{\text{T}\}]P[\{\text{H}\}] = \frac{1}{8}, \\
P[\{\text{THT}\}] &= P[\{\text{T}\}]P[\{\text{H}\}]P[\{\text{T}\}] = \frac{1}{8}, \\
P[\{\text{HTT}\}] &= P[\{\text{H}\}]P[\{\text{T}\}]P[\{\text{T}\}] = \frac{1}{8}, \text{ and} \\
P[\{\text{TTT}\}] &= P[\{\text{T}\}]P[\{\text{T}\}]P[\{\text{T}\}] = \frac{1}{8}.
\end{aligned}$$

Example 5: System Reliability

A system consists of a controller and three peripheral units. The system is said to be "up" if the controller and at least two of the peripherals are functioning. Find the probability that the system is up assuming that all components fail independently.

Define the following events: A is "controller is functioning"; and B_i is "peripheral i is functioning"; where $i = 1, 2, 3$. The event F, "two or more peripheral units are functioning," occurs if all three units are functioning or if exactly two units are functioning. Thus

$$\begin{aligned} F &= [B_1 \cap B_2 \cap B_3^c) \cup (B_1 \cap B_2^c \cap B_3) \\ &\cup (B_1^c \cap B_2 \cap B_3) \cup (B_1 \cap B_2 \cap B_3) \end{aligned}$$

Note that the events in the above union are mutually exclusive. Thus

$$\begin{aligned} P[F] =& P[B_1]\, P[B_2]\, P[B_3^c] + P[B_1]\, P[B_2^c]\, P[B_3] \\ & + P[B_1^c]\, P[B_2]\, P[B_3] + P[B_1]\, P[B_2]\, P[B_3]\} \\ =& 3(1-a)^2 a + (1-a)^3 \end{aligned}$$

where we have assumed that each peripheral fails with probability a, so that $P[B_i] = 1 - a$ and $P[B_i^c] = a$.

The event "system is up" is then $A \cap F$. If we assume that the controller fails with probability p, then

$$\begin{aligned} P[\text{ "system up"}] &= P[A \cap F] = P[A]P[F] \\ &= (1-p)P[F] \\ &= (1-p)\left\{3(1-a)^2 a + (1-a)^3\right\} \end{aligned}$$

Let $a = 10\%$, then all three peripherals are functioning $(1-a)^3 = 72.9\%$ of the time and two are functioning and one "down" $3(1-a)^2 a = 24.3\%$ of the time. Thus two or more peripherals are functioning 97.2% of the time. Suppose that the controller is not very reliable, say $p = 20\%$, then the system is up only 77.8% of the time, mostly because of controller failures.

Suppose a second identical controller with $p = 20\%$ is added to the system, and that the system is "up" if at least one of the controllers is functioning and if two or more of the peripherals are functioning. In Problem 65, you are asked to show that at least one of the controllers is functioning 96% of the time, and that the system is up 93.3% of the time. This is an increase of 16% over the system with a single controller.

2 Sequential Experiments

Many random experiments consist of a sequence of simpler subexperiments.

- These subexperiments may or may not be independent.
- In this section we discuss methods for obtaining the probabilities of events in sequential experiments.

2.1 Sequences of Independent Experiments

- A random experiment consists of performing *subexperiments* $E_1, E_2, \ldots, E_n$.
- The outcome of this experiment will then be an n-tuple $s = (s_1, \ldots, s_n)$, where s_k is the outcome of the kth subexperiment.
- The sample space of the sequential experiment is defined as the set that contains the above n-tuples and is denoted by the Cartesian product of the individual sample spaces $S_1 \times S_2 \times \cdots \times S_n$.
- We can usually determine, because of physical or logical considerations, when the subexperiments are independent, in the sense that the outcome of any given subexperiment cannot affect the outcomes of the other subexperiments.
- Let $A_1, A_2, \ldots, A_n$ be events such that A_k concerns only the outcomes of the kth subexperiment.
- If the subexperiments are independent, then it is reasonable to assume that the above events $A_1, A_2, \ldots, A_n$ are independent. Thus

  $$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]\, P[A_2] \ldots P[A_n]$$

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

Example 6

Suppose that 10 numbers are selected at random from the interval $[0, 1]$. Find the probability that the first 5 numbers are less than $1/4$ and the last 5 numbers are greater than $1/2$. Let $x_1, x_2, \ldots, x_{10}$ be the sequence of 10 numbers, then the events of interest are:

$$\begin{aligned} A_k &= \left\{x_k < \tfrac{1}{4}\right\} \quad \text{for } k = 1, \ldots, 5 \\ A_k &= \left\{x_k > \tfrac{1}{2}\right\} \quad \text{for } k = 6, \ldots, 10 \end{aligned}$$

If we assume that each selection of a number is independent of the other selections, then

$$\begin{aligned} P[A_1 \cap A_2 \cap \cdots \cap A_{10}] &= P[A_1]\, P[A_2] \ldots P[A_{10}] \\ &= \left(\frac{1}{4}\right)^5 \left(\frac{1}{2}\right)^5 \end{aligned}$$

- We will now derive several important models for experiments that consist of sequences of independent subexperiments.

2.2 The Binomial Probability Law

A **Bernoulli trial** involves performing a subexperiment once and noting whether a particular event A occurs.

- The outcome of the Bernoulli trial is said to be a "success" if A occurs and a "failure" otherwise.
- We are interested in finding the probability of k successes in n independent repetitions of a Bernoulli trial.
- We can view the outcome of a single Bernoulli trial as the outcome of a toss of a coin for which the probability of heads (success) is $p = P[A]$.
- The probability of k successes in n Bernoulli trials is then equal to the probability of k heads in n tosses of the coin.

Example 7

Suppose that a coin is tossed three times. If we assume that the tosses are independent and the probability of heads is p, then the probability for the sequences of heads and tails is

$$\begin{aligned} &P[\{\text{HHH}\}] = P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{H}\}] = p^3, \\ &P[\{\text{HHT}\}] = P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{T}\}] = p^2(1-p), \\ &P[\{\text{HTH}\}] = P[\{\text{H}\}]P[\{\text{T}\}]P[\{\text{H}\}] = p^2(1-p), \end{aligned}$$

$$
\begin{aligned}
&P[\{\mathrm{THH}\}] = P[\{\mathrm{T}\}]P[\{\mathrm{H}\}]P[\{\mathrm{H}\}] = p^2(1-p),\\
&P[\{\mathrm{TTH}\}] = P[\{\mathrm{T}\}]P[\{\mathrm{T}\}]P[\{\mathrm{H}\}] = p(1-p)^2,\\
&P[\{\mathrm{THT}\}] = P[\{T\}]P[\{\mathrm{H}\}]P[\{\mathrm{T}\}] = p(1-p)^2,\\
&P[\{\mathrm{HTT}\}] = P[\{\mathrm{H}\}]P[\{\mathrm{T}\}]P[\{\mathrm{T}\}] = p(1-p)^2, \text{ and}\\
&P[\{\mathrm{TTT}\}] = P[\{\mathrm{T}\}]P[\{\mathrm{T}\}]P[\{\mathrm{T}\}] = (1-p)^3,
\end{aligned}
$$

where we used the fact that the tosses are independent. Let k be the number of heads in three trials, then

$$
\begin{aligned}
&P[k=0] = P[\{\mathrm{TTT}\}] = (1-p)^3,\\
&P[k=1] = P[\{\mathrm{TTH}, \mathrm{THT}, \mathrm{HTT}\}] = 3p(1-p)^2,\\
&P[k=2] = P[\{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}\}] = 3p^2(1-p), \text{ and}\\
&P[k=3] = P[\{\mathrm{HHH}\}] = p^3.
\end{aligned}
$$

The result in Example 7 is the $n = 3$ case of the binomial probability law.

Theorem 2.1. *Let k be the number of successes in n independent Bernoulli trials, then the probabilities of k are given by the* ***binomial probability law****:*

$$
p_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \textit{for} \quad k = 0, \ldots, n, \tag{4}
$$

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

$$
\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{5}
$$

is the ***Binomial coefficient****.*

- The term $n!$ in Eq. (5) is called n factorial and is defined by $n! = n(n-1)\ldots(2)(1)$. By definition $0!$ is equal to 1.

Proof. In Example 7 we saw that each of the sequences with k successes and $n-k$ failures has the same probability, namely $p^k(1-p)^{n-k}$. Let $N_n(k)$ be the number of distinct sequences that have k successes and $n-k$ failures, then

$$
p_n(k) = N_n(k) p^k (1-p)^{n-k} \tag{6}
$$

The expression $N_n(k)$ is the number of ways of picking k positions out of n
for the successes. It can be shown that [2]

$$N_n(k) = \binom{n}{k}$$

The theorem follows by substituting Eq. (7) into Eq. (6). □

Example 8

Verify that Eq. (4) gives the probabilities found in Example 7.

In Example 7, let "toss results in heads" correspond to a "success," then

$$\begin{aligned}
p_3(0) &= \frac{3!}{0!3!}p^0(1-p)^3 = (1-p)^3, \\
p_3(1) &= \frac{3!}{1!2!}p^1(1-p)^2 = 3p(1-p)^2, \\
p_3(2) &= \frac{3!}{2!1!}p^2(1-p)^1 = 3p^2(1-p), \text{ and} \\
p_3(3) &= \frac{3!}{0!3!}p^3(1-p)^0 = p^3
\end{aligned}$$

which are in agreement with our previous results.

- The binomial coefficient is usually introduced in an introductory calculus course when the **binomial theorem** is discussed:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \tag{7}$$

If we let $a = b = 1$, then

$$2^n = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} N_n(k)$$

[2] See Example 4 in Week 2 Notes.

which is in agreement with the fact that there are 2^n distinct possible sequences of successes and failures in n trials. If we let $a = p$ and $b = 1 - p$ in Eq. (7) we then obtain

$$1 = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} p_n(k)$$

which confirms that the probabilities of the binomial probabilities sum to 1.

- $n!$ grows very quickly with n, so numerical problems are encountered for relatively small values of n if one attempts to compute $p_n(k)$ directly using Eq. (4).
- The following recursive formula extends the range of n for which $p_n(k)$ can be computed before encountering numerical difficulties:

$$p_n(k+1) = \frac{(n-k)p}{(k+1)(1-p)} p_n(k) \tag{8}$$

- Later in the course, we present the Gaussian and Poisson approximations for the binomial probabilities for the case when n is large.

Example 9

Let k be the number of active (nonsilent) speakers in a group of eight noninteracting (i.e., independent) speakers. Suppose that a speaker is active with probability 1/3. Find the probability that the number of active speakers is greater than six.

For $i = 1, \ldots, 8$, let A_i denote the event " ith speaker is active." The number of active speakers is then the number of successes in eight Bernoulli trials with $p = 1/3$. Thus the probability that more than six speakers are active is

$$\begin{aligned} P[k=7] + P[k=8] &= \binom{8}{7}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right) + \binom{8}{8}\left(\frac{1}{3}\right)^8 \\ &= .00244 + .00015 = .00259 \end{aligned}$$

Example 10: Error Correction Coding

A communication system transmits binary information over a channel that introduces random bit errors with probability $\varepsilon = 10^{-3}$. The transmitter transmits each information bit three times, and a decoder takes a majority vote of the received bits to decide on what the transmitted bit was. Find the probability that the receiver will make an incorrect decision.

The receiver will make the wrong decision if the channel introduces two or more errors. If we view each transmission as a Bernoulli trial in which a "success" corresponds to the introduction of an error, then the probability of two or more errors in three Bernoulli trials is

$$P[k \geq 2] = \binom{3}{2}(.001)^2(.999) + \binom{3}{3}(.001)^3 \simeq 3\left(10^{-6}\right)$$

2.3 The Multinomial Probability Law

The binomial probability law can be generalized to an experiment where we note the occurrence of more than one event.

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

  $$p_1 + p_2 + \cdots + p_M = 1$$

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

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

where $k_1 + k_2 + \cdots + k_M = n$.

The binomial probability law is the $M = 2$ case of the multinomial probability law.

Example 11

A dart is thrown nine times at a target consisting of three areas. Each throw has a probability of .2, .3, and .5 of landing in areas 1,2, and 3, respectively. Find the probability that exactly three darts land in each of the areas.

This experiment consists of nine independent repetitions of a subexperiment that has three possible outcomes. The probability for the number of occurrences of each outcome is given by the multinomial probabilities with parameters $n = 9$ and $p_1 = .2, p_2 = .3$, and $p_3 = .5$:

$$P[(3,3,3)] = \frac{9!}{3!3!3!}(.2)^3(.3)^3(.5)^3 = .04536$$

Example 12

Suppose we pick 10 telephone numbers at random from a telephone book and note the last digit in each of the numbers. What is the probability that we obtain each of the integers from 0 to 9 only once?

The probabilities for the number of occurrences of the integers is given by the multinomial probabilities with parameters $M = 10, n = 10$, and $p_j = 1/10$ if we assume that the 10 integers in the range 0 to 9 are equiprobable. The probability of obtaining each integer once in 10 draws is then

$$\frac{10!}{1!1!\ldots 1!}(.1)^{10} \simeq 3.6\left(10^{-4}\right)$$

2.4 The Geometric Probability Law

In a sequential experiment we repeat independent Bernoulli trials until the occurrence of the first success. Let the outcome be m, the number of trials carried out until the occurrence of the first success.

- The sample space for this experiment is the set of positive integers.
- The probability, $p(m)$, that m trials are required is found by noting that this can only happen if the first $m - 1$ trials result in failures

and the mth trial in success.[4] The probability of this event is

$$p(m) = P\left[A_1^c A_2^c \ldots A_{m-1}^c A_m\right] = (1-p)^{m-1}p \quad m = 1, 2, \ldots \tag{9}$$

where A_i is the event "success in ith trial." The probability assignment specified by Eq.(9) is called the geometric probability law.

- The probabilities in Eq. (9) sum 1:

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

where $q = 1 - p$, and where we have used the formula for the summation of a geometric series.

- The probability that more than K trials are required before a success occurs has a simple form:

$$\begin{aligned} P[\{m > K\}] &= p \sum_{m=K+1}^{\infty} q^{m-1} = pq^K \sum_{j=0}^{\infty} q^j \\ &= pq^K \frac{1}{1-q} \\ &= q^K. \end{aligned} \tag{10}$$

Example 13: Error Control by Retransmission

Computer A sends a message to computer B over an unreliable telephone line. The message is encoded so that B can detect when errors have been introduced into the message during transmission. If B detects an error it requests A to retransmit it. If the probability of a message transmission error is $q = .1$, what is the probability that a message needs to be transmitted more than two times?

Each transmission of a message is a Bernoulli trial with probability of success $p = 1 - q$. The Bernoulli trials are repeated until the first success (error-free transmission). The probability that more than two transmissions are required is given by Eq. (10):

$$P[m > 2] = q^2 = 10^{-2}.$$

2.5 Sequences of Dependent Experiments

We now consider a sequence or "chain" of subexperiments in which the outcome of a given subexperiment determines which subexperiment is performed next.

Example 14

A sequential experiment involves repeatedly drawing a ball from one of two urns, noting the number on the ball, and replacing the ball in its urn. Urn 0 contains a ball with the number 1 and two balls with the number 0, and urn 1 contains five balls with the number 0 and one ball with the number 1. The urn from which the first draw is made is selected at random by flipping a fair coin. Urn 0 is used if the outcome is heads and urn 1 if the outcome is tails. Thereafter the urn used in a subexperiment corresponds to the number on the ball selected in the previous subexperiment.

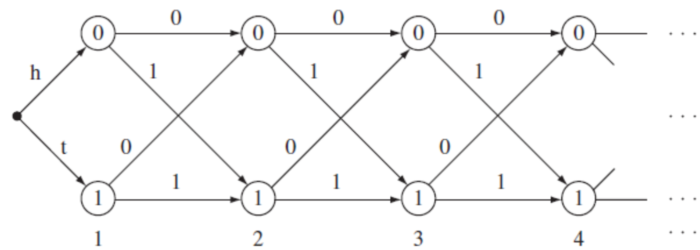
The sample space of this experiment consists of sequences of 0s and 1s. Each possible sequence corresponds to a path through the "trellis" diagram shown in Fig. 3(a). The nodes in the diagram denote the urn used in the nth subexperiment, and the labels in the branches denote the outcome of a subexperiment. Thus the path 0011 corresponds to the sequence: The coin toss was heads so the first draw was from urn 0; the outcome of the first draw was 0, so the second draw was from urn 0; the outcome of second draw was 1, so the third draw was from urn 1; and the outcome from the third draw was 1, so the fourth draw is from urn 1.

Now suppose that we want to compute the probability of a particular sequence of outcomes, say s_0, s_1, s_2. Denote this probability by $P\left[\{s_0\} \cap \{s_1\} \cap \{s_2\}\right]$. Let $A = \{s_2\}$ and $B = \{s_0\} \cap \{s_1\}$, then since $P[A \cap B] = P[A \mid B]P[B]$ we have

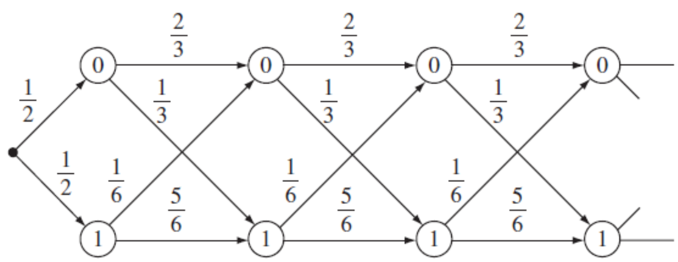
$$\begin{aligned} P\left[\{s_0\} \cap \{s_1\} \cap \{s_2\}\right] &= P\left[\{s_2\} \mid \{s_0\} \cap \{s_1\}\right] P\left[\{s_0\} \cap \{s_1\}\right] \\ &= P\left[\{s_2\} \mid \{s_0\} \cap \{s_1\}\right] P\left[\{s_1\} \mid \{s_0\}\right] P\left[\{s_0\}\right] \end{aligned}$$

Note that in the above urn example the probability $P[\{s_n\} \mid \{s_0\} \cap \cdots \cap \{s_{n-1}\}]$ depends only on $\{s_{n-1}\}$ since the most recent outcome determines which subexperiment is performed:

$$P\left[\{s_n\} \mid \{s_0\} \cap \cdots \cap \{s_{n-1}\}\right] = P\left[\{s_n\} \mid \{s_{n-1}\}\right] \tag{11}$$



(a) Each sequence of outcomes corresponds to a path through this trellis diagram.



(b) The probability of a sequence of outcomes is the product of the probabilities along the associated path.

Figure 3: Trellis Diagram for a Markov Chain

Therefore for the sequence of interest we have that

$$P[\{s_0\} \cap \{s_1\} \cap \{s_2\}] = P[\{s_2\} \mid \{s_1\}]\, P[\{s_1\} \mid \{s_0\}]\, P[\{s_0\}]$$

Sequential experiments that satisfy Eq. (11) are called **Markov chains**.

Example 15

Find the probability of the sequence 0011 for the urn experiment introduced in previous example.

Recall that urn 0 contains two balls with label 0 and one ball with label 1, and that urn 1 contains five balls with label 1 and one ball with label 0. We can readily compute the probabilities of sequences of outcomes by labeling the branches in the trellis diagram with the probability of the corresponding transition as shown in Fig. 3(b). Thus the probability of the sequence 0011 is given by

$$P[0011] = P[1 \mid 1]P[1 \mid 0]P[0 \mid 0]P[0]$$

where the transition probabilities are given by

$$\begin{array}{lcl} P[1 \mid 0] = \frac{1}{3} & \text{and} & P[0 \mid 0] = \frac{2}{3} \\ P[1 \mid 1] = \frac{5}{6} & \text{and} & P[0 \mid 1] = \frac{1}{6} \end{array}$$

and the initial probabilities are given by

$$P(0) = \frac{1}{2} = P[1]$$

If we substitute these values into the expression for $P[0011]$, we obtain

$$P[0011] = \left(\frac{5}{6}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{5}{54}.$$

- The outcomes in many random experiments are numbers or vectors of numbers and so the events are subsets R^n.
- We now introduce the notion of a random variable which leads to methods that are useful in computing the probabilities of events in these types of random experiments.

3 The Notion of a Random Variable

The outcome of a random experiment need not be a number. However, we are usually interested not in the outcome itself, but rather in some measurement or numerical attribute of the outcome, e.g., the height of a person.

- For example, in n tosses of a coin, we may be interested in the total number of heads and not in the specific order in which heads and tails occur.
- In a randomly selected computer job, we may be interested only in the execution time of the job.
- In the selection of a student's name from an urn, we may be interested in the weight of the student.

In each of these examples, a measurement assigns a numerical value to the outcome of the random experiment. Since the outcomes are random, the results of the measurements will also be random and we are interested in the probabilities of the resulting numerical values.

A **random variable** X is a function that assigns a real number, $X(\zeta)$, to each outcome ζ in the sample space of a random experiment, as shown pictorially in Fig. 4.

- The specification of a measurement on the outcome of a random experiment defines a function on the sample space, and hence a random variable.

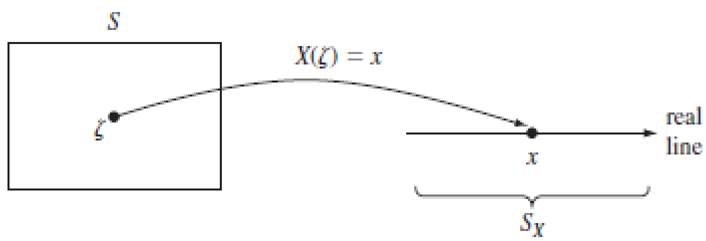


Figure 4: A random variable assigns a number $X(\zeta)$ to each outcome ζ in the sample space S of a random experiment.

Example 16

A coin is tossed three times and the sequence of heads and tails is noted. The sample space for this experiment is $S = \{$ hhh, hht, hth, htt, thh, tht, tth, ttt$\}$. Now let X be the number of heads in three coin tosses. X assigns each outcome ζ in S a number from the set $S_X = \{0, 1, 2, 3\}$. The table below lists the eight outcomes of S and the corresponding values of X.

ζ :	hhh	hht	hth	thh	htt	tht	tth	ttt
$X(\zeta)$:	3	2	2	2	1	1	1	0

X is then a random variable taking on values in the set $S_X = \{0, 1, 2, 3\}$.

- If the outcome ζ of some experiment is already a numerical value, we can immediately reinterpret the outcome as a random variable defined by the identity function, $X(\zeta) = \zeta$.

- Thus many of the examples considered in the previous chapter can now be viewed as random variables.

- We note that the function or rule that assigns values to each outcome is fixed and deterministic, as for example, in the rule "count the number of heads in three tosses of a coin."

- The randomness in the observed values is due to the underlying randomness of the arguments of the function X, namely the experiment outcomes ζ.

- In other words, the randomness in the observed values of X are *induced* by the underlying random experiment, and we should therefore be able to compute the probabilities of the observed values in terms of the probabilities of the underlying outcomes.

Example 17

The event $\{X = k\} = \{k \text{ heads in three coin tosses }\}$ occurs when the outcome of the coin tossing experiment contains k heads. The probability of the event $\{X = k\}$ is therefore given by the sum of the probabilities of the corresponding outcomes or elementary events. In Example 4, we found the probabilities of the elementary events of the coin tossing experiment. Thus we have

$$\begin{aligned}
p_0 &= P[X = 0] = P[\{\text{ ttt }\}] = (1-p)^3 \\
p_1 &= P[X = 1] = P[\{\text{ htt }\}] + P[\{\text{ tht }\}] + P[\{\text{ tth }\}] = 3(1-p)^2 p \\
p_2 &= P[X = 2] = P[\{\text{ hht }\}] + P[\{\text{ hth }\}] + P[\{\text{ thh }\}] = 3(1-p)p^2
\end{aligned}$$

and

$$p_3 = P[X = 3] = P[\{\text{hhh}\}] = p^3$$

The p_k's can be used to obtain the probabilities of all events that involve X.

- *If we are concerned only with the values of X, we can ignore the underlying experiment with sample space S, and proceed as if the experiment consisted of sample space S_X with probabilities p_k.*

Example 17 demonstrates the following general technique using equivalent events for finding the probabilities of events involving the random variable X.

- Let S_X be the set of values that can be taken on by X, and let B be some subset of S_X. S_X can be viewed as a new sample space, and B as an event in the sample space.
- Let A be the set of outcomes ζ in S that lead to values $X(\zeta)$ in B, as shown in Fig. 5, that is, $A = \{\zeta : X(\zeta) \text{ in } B\}$ then the event B in S_X occurs whenever the event A in S occurs.
- Thus the probability of event B is given by

$$P[B] = P[A] = P[\{\zeta : X(\xi) \text{ in } B\}].$$

- We refer to events A and B as **equivalent events**.

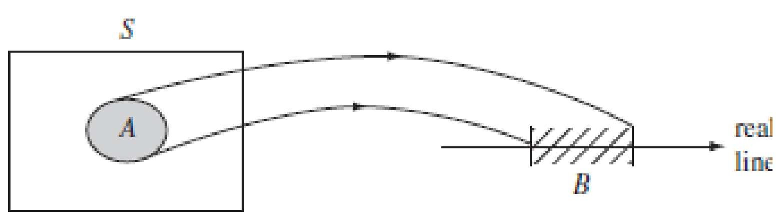


Figure 5: $P[X \text{ in } B] = P[\zeta \text{ in } A]$

4 The Cumulative Distribution Function

The **cumulative distribution function** (cdf) of a random variable X is defined as the probability of the event $\{X \leq x\}$:

$$F_X(x) = P[X \leq x] \quad \text{for } -\infty < x < +\infty \tag{12}$$

that is, the cdf is the probability that the random variable X takes on a value in the set $(-\infty, x]$.

- In terms of the underlying sample space, the cdf is the probability of the event $\{\zeta : X(\zeta) \leq x\}$.
- The event $\{X \leq x\}$ and its probability vary as x is varied; in other words $F_X(x)$ is a function of the variable x.
- The cdf has the following interpretation in terms of relative frequency. Suppose that the experiment that yields the outcome ζ, and hence $X(\zeta)$, is performed a large number of times. $F_X(b)$ is then the long-term proportion of times in which $X(\zeta) \leq b$.

The cdf is simply a convenient way of specifying the probability of all semi-infinite intervals of the real line of the form $(-\infty, x]$. The events of interest when dealing with numbers are intervals of the real line, and their complements, unions, and intersections.

- We show below that the probabilities of all of these events can be expressed in terms of the cdf.

The axioms of probability and their corollaries imply that the cdf has the following properties:

i. $0 \leq F_X(x) \leq 1$
ii. $\lim_{x\to\infty} F_X(x) = 1$
iii. $\lim_{x\to-\infty} F_X(x) = 0$.
iv. $F_X(x)$ is a nondecreasing function of x, that is, if $a < b$, then $F_X(a) \leq F_X(b)$.
v. $F_X(x)$ is continuous from the right, that is, for $h > 0$ $F_X(b) = \lim_{h\to 0} F_X(b+h) = F_X(b^+)$.

- The first property follows from the fact that the cdf is a probability and hence must satisfy Axiom I and Corollary 2.

- The second property follows from the fact that the event $\{X < \infty\}$ consists of all the real numbers and is thus the entire sample space (Axiom II).

- The third property follows from the fact that all real numbers are greater than $-\infty$, so the event $\{X \leq -\infty\}$ is the empty set (Corollary 3).

- To obtain the fourth property, note that the event $\{X \leq a\}$ is a subset of $\{X \leq b\}$ so it must have smaller or equal probability (Corollary 7).

- We will see how the fifth property comes about in Example 18.[3]

The probability of events that correspond to intervals of the form $\{a < X \leq b\}$ *can be expressed in terms of the cdf:*

$$\text{vi.} \quad P[a < X \leq b] = F_X(b) - F_X(a). \tag{13}$$

To show Eq. (13) we note that since

$$\{X \leq a\} \cup \{a < X \leq b\} = \{X \leq b\}$$

and since the two events on the left-hand side are mutually exclusive, we have by Axiom III and by Eq. (12) that

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

Equation (13) then follows.

Equation (13) allows us to compute the probability of the event $\{X = b\}$. Let $a = b - \varepsilon$ in Eq. (13), $\varepsilon > 0$, then

$$P[b - \varepsilon < X \leq b] = F_X(b) - F_X(b - \varepsilon)$$

As $\varepsilon \to 0$, the left side of the above equation approaches $P[X = b]$, thus

$$\text{vii.} \quad P[X = b] = F_X(b) - F_X\left(b^-\right). \tag{14}$$

[3] The proof that the cdf is continuous on the right is beyond the level intended here. The proof can be found in Leon-Garcia 3E.

Thus the probability that an arbitrary random variable X takes on a specific value, say b, is given by the magnitude of the jump of the cdf at the point b. It follows that if the cdf is continuous at a point b, then the event $\{X = b\}$ has probability zero.

Equation (14) can be combined with Eq. (13) to compute the probabilities of other types of intervals. For example, since

$$\{a \leq X \leq b\} = \{X = a\} \cup \{a < X \leq b\}$$

then

$$\begin{aligned} P[a \leq X \leq b] &= P[X = a] + P[a < X \leq b] \\ &= F_X(a) - F_X\left(a^-\right) + F_X(b) - F_X(a) \\ &= F_X(b) - F_X\left(a^-\right) \end{aligned} \tag{15}$$

Note that if the cdf is continuous at the endpoints of an interval, then the endpoints have zero probability, and therefore they can be included in, or excluded from, the interval without affecting the probability. In other words, *if the cdf is continuous at the points $x = a$ and $x = b$, then* the following probabilities are equal:

$$P[a < X < b], \quad P[a \leq X < b], \quad P[a < X \leq b], \text{ and } P[a \leq X \leq b]$$

Example 18

Figure 6(a) shows the cdf of the random variable X, which is defined as the number of heads in three tosses of a fair coin. From Example 16 we know that X takes on only the values $0, 1, 2$, and 3 with probabilities 1/8, $3/8, 3/8$, and 1/8, respectively, so $F_X(x)$ is simply the sum of the probabilities of the outcomes from $\{0, 1, 2, 3\}$ that are less than or equal to x. The resulting cdf is seen to have discontinuities at the points $0, 1, 2, 3$.

Let us take a closer look at one of these discontinuities. Consider the cdf in the vicinity of the point $x = 1$. For δ a small positive number, we have

$$F_X(1 - \delta) = P[X \leq 1 - \delta] = P\{0 \text{ heads }\} = \frac{1}{8},$$

so the limit of the cdf as x approaches 1 from the left is 1/8. However,

$$F_X(1) = P[X \leq 1] = P[0 \text{ or } 1 \text{ heads }] = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$$

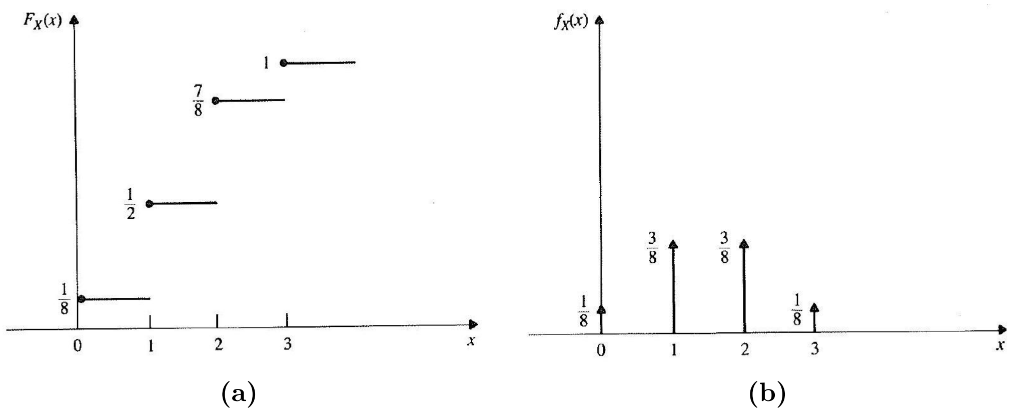


Figure 6: An example of a discrete random variable-the binomial random variable, $n = 3, p = 1/2$, showing its cdf (a) and pdf (b).

and also

$$F_X(1+\delta) = P[X \leq 1+\delta] = P[0 \text{ or } 1 \text{ heads }] = \frac{1}{2}.$$

Thus the cdf is continuous from the right and equal to 1/2 at the point $x = 1$. Indeed we note the magnitude of the jump at the point $x = 1$ is equal to $P[X = 1] = 1/2 - 1/8 = 3/8$.

The cdf can be written compactly in terms of the unit step function:

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

then

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

Example 19: Exponential Random Variable

The transmission time X of messages in a communication system obey the exponential probability law with parameter λ, that is,

$$P[X > x] = e^{-\lambda x} \quad x > 0$$

Find the cdf of X. Find $P[T < X \leq 2T]$, where $T = 1/\lambda$.

The cdf of X is $F_X(x) = P[X \le x] = 1 - P[X > x]$:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

The cdf is shown in Fig. 7(a). From property (vi) we have

$$P[T < X \le 2T] = 1 - e^{-2} - \left(1 - e^{-1}\right) = e^{-1} - e^{-2} \simeq .233$$

Note that $F_X(x)$ is continuous for all x. Note also that its derivative exists everywhere except at $x = 0$:

$$F'_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases}$$

$F'(x)$ is shown in Fig. 7(b).

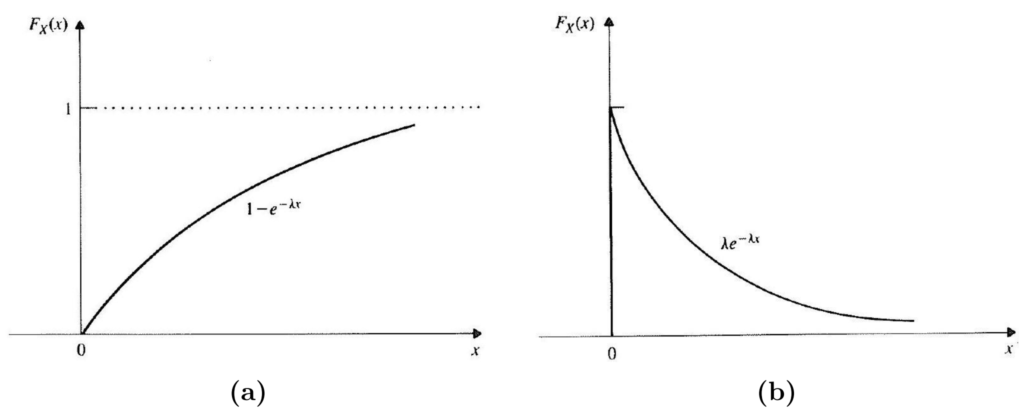


Figure 7: An example of a continuous random variable-the exponential random variable, its cdf (a) and pdf (b).

Example 20

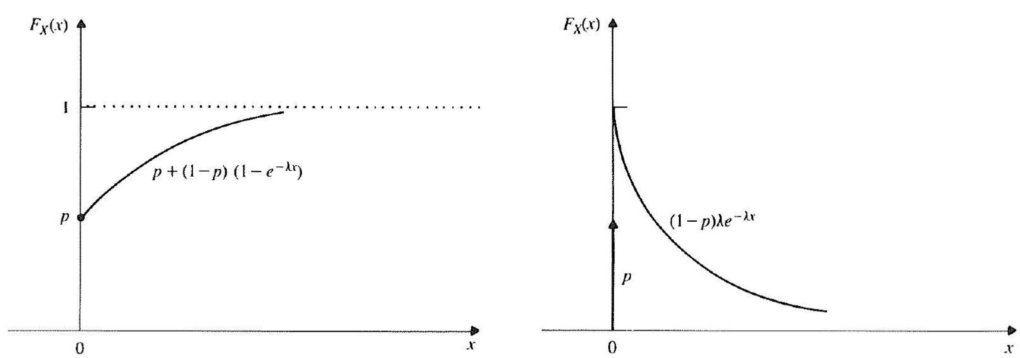
The waiting time W of a customer in a queueing system is zero if he finds the system idle, and an exponentially distributed random length of time if he finds the system busy. The probabilities that he finds the system idle or busy are p and $1-p$, respectively. Find the cdf of W.

The cdf of W is found as follows:

$$\begin{aligned} F_X(x) &= P[W \leq x] \\ &= P[W \leq x \mid \text{ idle }]p + P[W \leq x \mid \text{ busy }](1-p) \end{aligned}$$

where the last equality used the theorem of total probability (from Week 2 Notes). Noting that $P[W \leq x \mid \text{idle}] = 1$ when $x \geq 0$ and 0 otherwise, we have

$$F_X(x) = \begin{cases} 0 & x < 0 \\ p + (1-p)\left(1 - e^{-\lambda x}\right) & x \geq 0 \end{cases}$$



(a) **(b)**

Figure 8: An example of a random variable of mixed type, its cdf (a) and pdf (b).

The cdf is shown in Fig. 8(a). Note that $F_X(x)$ can be expressed as the sum of a step function with amplitude p and a continuous function of x.

4.1 The Three Types of Random Variables

The random variables in Examples 18, 19, and 20 are typical of the three basic types of random variable that we will be interested in.

A **discrete random variable** is defined as a random variable whose cdf is a right-continuous, staircase function of x, with jumps at a countable set of points $x_0, x_1, x_2, \ldots$ The random variable in Example 18 is an example of a discrete random variable. Discrete random variables take on values from a finite or at most a countably infinite set $S_X = \{x_0, x_1, \ldots\}$. They arise mostly in applications that involve counting, so we usually have $S_X = \{0, 1, 2, \ldots\}$.

The cdf of a discrete random variable can be written as the weighted sum of unit step functions as in Example 18:

$$F_X(x) = \sum_k p_X(x_k)\, u(x - x_k)$$

where $p_X(x_k) = P[X = x_k]$ gives the magnitude of the jumps in the cdf. The set of probabilities $p_X(x_k)$ is called the probability mass function (pmf) of X.

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

$$F_X(x) = \int_{-\infty}^{x} f(t)dt \tag{16}$$

For **continuous random variables**, the cdf is continuous everywhere, so property (vii) implies that $P[X = x] = 0$ for all x. The cdf of the random variable discussed in Example 19 is a continuous random variable since its cdf is continuous everywhere, and since Eq. (16) is satisfied if we let $f(x) = F'_X(x)$ as given in the example.

A **random variable of mixed type** is a random variable with a cdf that has jumps on a countable set of points $x_0, x_1, x_2, \ldots$ but that also increases continuously over at least one interval of values of x. The cdf for these random variables has the form

$$F_X(x) = pF_1(x) + (1 - p)F_2(x)$$

where $0 < p < 1$, and $F_1(x)$ is the cdf of a discrete random variable and $F_2(x)$ is the cdf of a continuous random variable. The random variable in Example 20 is of mixed type.

Random variables of mixed type can be viewed as being produced by a two-step process: A coin is tossed; if the outcome of the toss is heads, a discrete random variable is generated according to $F_1(x)$, otherwise, a continuous random variable is generated according to $F_2(x)$.