Chapter 3

Continuous Random Variables

3.1 Introduction

In the previous chapter, we described a random variable X as a measurable function from a probability space $(\Omega, \mathcal{E}, \mathbb{P})$ to the real line $\Re$. We then focused on studying discrete random variables where the range of X, denoted by $R_X = X(\Omega)$, has a discrete, possibly countably infinite number of elements. However, what about random variables where the range of X, denoted by R_X has an uncountable number of elements? This is illustrated in Figure 3.1 where the range of X maps Ω into an interval $[a, b]$. Suppose we wanted to define a uniform probability on that interval $[a, b]$. In this case, it is impossible to assign a probability mass to any point $x \in [a, b]$, other than 0, because we could not satisfy both the additivity property (the probability of the union of disjoint sets is the sum of the probabilities of the individual sets) and the normalization property (the probability of that $X(\omega) \in R_X$ equals 1).

In cases where the range R_X is uncountable, it is common that one cannot associate a nonzero probability with any individual outcome. Since there are uncountably many values of the random variable $X(\omega), \omega \in S$, we focus on defining probabilities of events, and not individual outcomes. In terms of events, our focus will be on events generated by the random variable X taking values in Borel sets: sets generated by countable unions, complements and intersections of intervals. By restricting the random variable X to be measurable, we guarantee that the inverse image of such a Borel set B, $\{\omega \in \Omega : X(\omega) \in B\}$ is an event in the event space $\mathcal{E}$, and thus has a probability assigned to it by the measure $\mathbb{P}$.

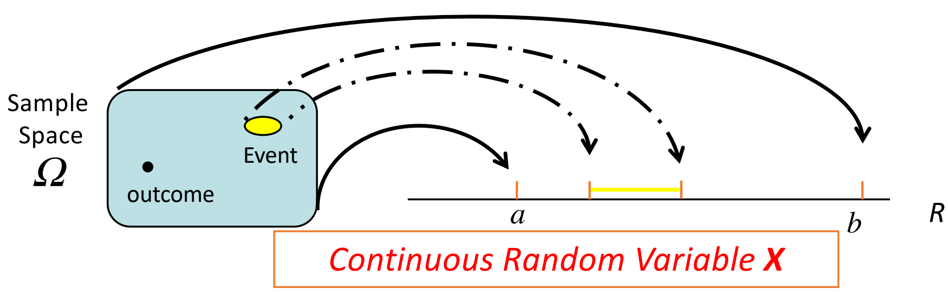

Figure 3.1: A continuous random variable has an uncountable range.

3.2 Continuous Random Variables

In Chapter ??, section ??, we defined the cumulative distribution function of a random variable X as:

$$F_X(a) \equiv \mathbb{P}_X(\{X \in (-\infty, a\}]) = \mathbb{P}[\{\omega \in \Omega : X(\omega) \leq a\}].$$

This definition is valid for all random variables, independent of whether the range R_X is discrete or not. The function $F_X(a)$ is defined for all $a \in \Re$.

This cumulative distribution function had the following properties:

1. (**Non-negativity**) $F_X(x) \geq 0$.

2. (**Normalization**) $F_X(\infty) = 1, F_X(-\infty) = 0$
3. (**Monotonicity**) $a \leq b$ implies that $F_X(a) \leq F_X(b)$
4. (**Right-continuity**) $\lim_{\epsilon \to 0^+} F_X(a + \epsilon) = F_X(a)$ (continuity from the right)

3.2.1 Cumulative Distribution Function

We will use the cumulative distribution of X to define a **continuous random variable**, although we will wait for a more precise definition later. Unlike discrete random variables, a **continuous random variable** must have a continuous cumulative distribution function (CDF) $F_X(x)$, as illustrated in Figure 3.2. Discontinuities in CDFs occur at values x which occur with positive probability, so that $\mathbb{P}[\{\omega \in \Omega : X(\omega) = x\}] > 0$. For continuous random variables, we want the probability $\mathbb{P}[\{\omega \in \Omega : X(\omega) = x\}] = 0$ for all $x \in \Re$. Hence, the CDF must be continuous.

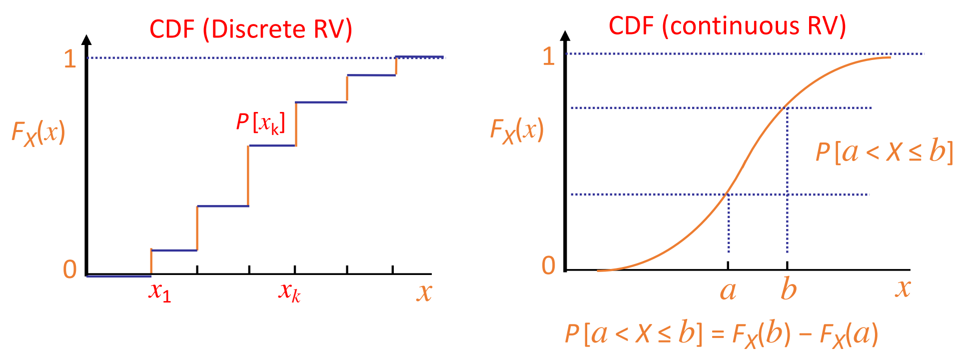

Figure 3.2: CDFs of discrete and continuous random variables.

The CDF of continuous random variables has the following additional properties:

1. **Continuity** $F_X(x)$ is a continuous function of x, i.e., $F_X(x) = \lim_{\epsilon \to 0} F_X(x + \epsilon)$.
2. $\mathbb{P}[\{\omega \in \Omega : X(\omega) = x\}] = 0$ for all $x \in \Re$. Every atom in R_X has zero probability.
3. $\mathbb{P}[\{\omega \in \Omega : X(\omega) \leq x\}] = \mathbb{P}[\{s : X(\omega) < x\}]$.
4. For $a < b, \mathbb{P}[\{\omega \in \Omega : a < X(\omega) \leq b\}] = F_X(b) - F_X(a)$.
5. If y is any number in the range $0 < y < 1$, then there must be at least one number x such that $F_X(x) = y$. This is a consequence of the intermediate value theorem for continuous functions. Note that there could be multiple such numbers, as illustrated in Figure 3.3

Example 3.1
Suppose we want to choose a random number in the interval $(0, 1)$, with every number equally likely to be chosen. That is, $R_X = (0, 1)$. Intuitively, the meaning of random in this instance is that we do not favor any one number over others in the interval (0,1). One way of expressing the innate randomness of the choice is as follows: Given any subinterval of (0, 1), the probability that the chosen number lies in that subinterval is equal to the length of that interval. One way of capturing this is with the following CDF $F_X(x)$:

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & x \in (0,1) \\ 1 & x \geq 1. \end{cases}$$

This cumulative distribution is illustrated in Figure 3.4. Note that the function is continuous; however, it is not differentiable at either $x = 0$ or $x = 1$, as the slopes from the left and right at those two points do not match.

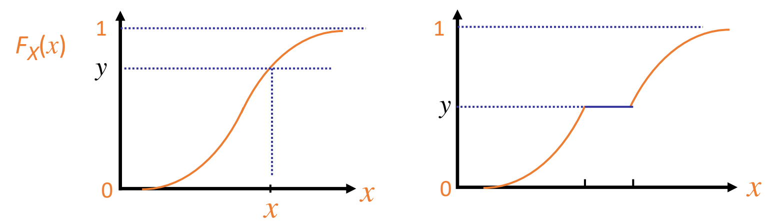

Figure 3.3: CDF where only one x satisfies $F_X(x) = y$, and where an interval of x satisfies $F_X(x) = y$.

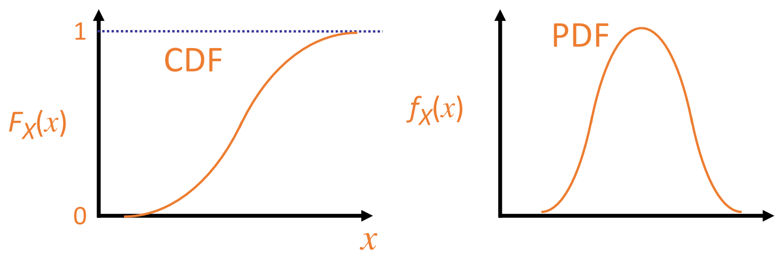

Figure 3.4: CDF and PDF for a continuous random variable.

For a random variable X to be continuous, it is not sufficient to have a continuous CDF. We want the CDF to be differentiable almost everywhere.[1] Formally, we define a continuous random variable below:

Definition 3.1
A random variable X is a **continuous random variable** if its cumulative distribution function $F_X(x)$ is continuous and differentiable almost everywhere. That is, its CDF can be written as an integral $F_X(x) = \int_{-\infty}^{x} f_X(x')d\,x'$ for some non-negative function $f_X(x')$. We refer to the function $f_X(x)$ as the **probability density function (PDF)**.

For a function to be differentiable almost everywhere, it must be differentiable everywhere except for a countable number of points $x_1, x_2, \ldots$, and there can only be a finite number of non-differentiable points in any finite-length interval. This means the CDF will have a probability density function:

$$f_X(x) = \begin{cases} \frac{d}{dx} F_X(x) & \text{if } F_X(x) \text{ is differentiable at } x, \\ \text{any non-negative number} & \text{otherwise.} \end{cases}$$

Figure 3.4 illustrates a cumulative distribution function for a continuous random variable and its corresponding probability density function (PDF). Note that this cumulative distribution function is differentiable everywhere, so the PDF is uniquely defined everywhere.

3.2.2 Probability Density Function

The PDF of a continuous random variable is not a probability and may take values greater than one, but it must be non-negative: It is a probability density. It is measured in units of probability per unit length. However, the **integral of a PDF over a region of** x is a probability, and must be a number in [0,1]. At this point, let's compare the concept of a PDF to the concept of a mass density for physical objects. Table 3.1 shows this comparison.

The probability density function for continuous random variables plays a similar role to the probability mass function for discrete random variables. The sum of the PMF of a discrete random variable over all the

[1] If you are curious, there are random variables with continuous CDF that are not differentiable almost everywhere. Look up references to Cantor distributions or the Devil's staircase function.

Physical mass in a system	probability in an experiment
Is non-negative	is non-negative
Density a function of space $\rho(x)$	probability density a function over the reals $\rho(x)$
Mass of region=	Probability of events =
Integral of density over region	Integral of density over outcomes in event

Table 3.1: Comparison of physical density and probability density

values in its domain R_X is equal to 1. Similarly, for a continuous random variable, the integral of its PDF over the entire real line is equal to 1. Although it is not a probability, if $f_X(a)$ is finite, then, for small ϵ, the probability that a sample value occurs in the interval $[a, a+\epsilon]$ is approximately $p_X(a)\epsilon$. Note that, as ϵ decreases, the probability that $X = a$ becomes zero.

The PDF satisfies the following basic properties:

1. **Non-negativity:** $f_X(x) \geq 0$.

   This follows from the monotonicity property of the cumulative distribution function, which is non-decreasing. Hence, its derivative, whenever it exists, is defined as

   $$f_X(x) = \lim_{\epsilon \to 0} \frac{F_X(x+\epsilon) - F_X(x)}{\epsilon} \ .$$

   For $\epsilon > 0$, the numerator inside the limit is always non-negative, and hence the limit, if it exists, must also be non-negative.

2. **Normalization:** $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

   By definition, we know $F_X(x) = \int_{-\infty}^{x} f_X(u)d\ u$. We also know, by the normalization property of CDFs, that $\lim_{x\to\infty} F_X(x) = 1$. Thus,

   $$\lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} \int_{-\infty}^{x} f_X(u)d\ u = \int_{-\infty}^{x} \infty f_X(u)d\ u = 1 \ .$$

3. **Probability of an interval:** $\mathbb{P}_X[\{a < X \leq b\}] = \displaystyle\int_a^b f_X(x)\,dx$.

   Since $\mathbb{P}_X[\{X = x\}] = 0$ for any $x \in \Re$, we have

   $$\mathbb{P}_X[\{a < X \leq b\}] = \mathbb{P}_X[\{a \leq X \leq b\}] = \mathbb{P}_X[\{a < X < b\}]$$

   From the CDF properties, we know

   $$\mathbb{P}_X[\{a \leq X \leq b\}] = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)d\ x - \int_{-\infty}^{a} f_X(x)d\ x = \int_a^b f_X(x)d\ x \ .$$

4. $\lim_{x\to\infty} f_X(x) = 0; \lim_{x\to-\infty} f_X(x) = 0$.

   As the magnitude of x gets large, the PDF curve must decay to zero. Otherwise, the integral of the PDF would keep growing unbounded as $|x|$ increased. Furthermore, the slope of the pdf must also decay to zero as $|x|$ grows unbounded.

5. **PDF $\to$ CDF:** $\displaystyle\int_{-\infty}^{x} f_X(u)\,du = F_X(x)$.

   This is the definition of the PDF.

Example 3.2

Consider a continuous random variable X, with PDF specified as

$$f_X(x) = \begin{cases} 3x^2 & x \in [0,1], \\ 0 & \text{elsewhere.}. \end{cases}$$

We note this satisfies the properties that we want in a PDF: It is non-negative, and it integrates to 1:

$$\int_{-\infty}^{\infty} f_X(x)d\,x = \int_0^1 3x^2 d\,x = 1\ .$$

Note that $f_X(1) = 3 > 1$ as it does not have to satisfy the bound for a probability. What this PDF indicates is that it is four times denser around $x = 1$ than around $x = 0.5$. If you generated independent samples of this random variable, the number of samples around 1 would be 4 times the number of samples around 0.5.

For discrete random variables X, the PMF provided the complete characterization of the probability properties of X. A similar property exists for continuous random variables X: The PDF provides the complete characterization its probability properties that we need for computing probabilities on the outcomes in $\Re$.

Example 3.3
A continuous random variable X has PDF

$$f_X(x) = \begin{cases} 0.75(1-x^2) & -1 \le x \le 1, \\ 0 & \text{otherwise..} \end{cases}$$

This density is illustrated in Figure 3.2.2. Compute $\mathbb{P}_X[\{0.25 \le X \le 1.25\}]$. Using the basic properties of the PDF, we know $\mathbb{P}_X[\{0.25 \le X \le 1.25\}] = \int_{0.25}^{1.25} f_X(x)dx$. However, note that the region of integration involves two different pieces of the definition of f_X. Hence,

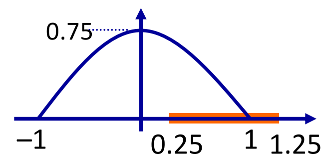

Figure 3.5: Figure for example 3.3.

$$\mathbb{P}_X[\{0.25 \le X \le 1.25\}] = \int_{0.25}^{1.25} f_X(x)dx = \int_{0.25}^{1} 0.75(1-x^2)dx + \int_1^{1.25} 0dx = \frac{81}{256}$$

Example 3.4
A continuous random variable X has PDF

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

This density is illustrated in Figure 3.2.2. Compute $F_X(-0.6)$ and $F_X(-0.3)$.

$$F_X(-0.6) = \int_{-\infty}^{-0.6} f_X(x)d\,x = \int_{-1}^{-0.6} (-2x)d\,x = 1 - 0.36 = 0.64\ .$$

$$F_X(-0.3) = \int_{-\infty}^{-0.3} f_X(x)d\,x = \int_{-1}^{-0.3} (-2x)d\,x = 1 - 0.09 = 0.91\ .$$

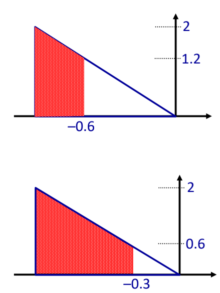

Figure 3.6: Figure for example 3.4.

Example 3.5
Assume a continuous random variable has a PDF given by

$$f_X(x) = \begin{cases} 0.75(1-x^2) & -1 \le x \le 1, \\ 0 & \text{otherwise..} \end{cases}$$

This density is illustrated in Figure 3.7. Compute $F_X(0)$ and $F_X(0.5)$.

$$F_X(0) = \int_{-\infty}^{0} f_X(x)d\,x = \frac{1}{2}\text{by symmetry!}$$

$$F_X(0.5) = \int_{-\infty}^{0.5} 0.75(1-x^2)d\,x = \int_{-1}^{0.5} 0.75(1-x^2)d\,x$$
$$= 0.75(1.5) - 0.25x^3|_{-1}^{0.5} = \frac{9}{8} - \frac{1}{4}(1+\frac{1}{8}) = \frac{27}{32}\ .$$

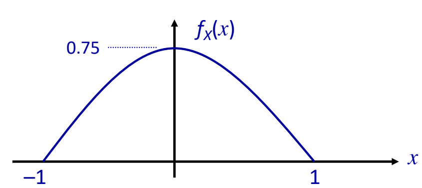

Figure 3.7: Figure for example 3.5.

3.3 Statistics of Continuous Random Variables

3.3.1 Expected Value

We have left the PDF undefined at points x where the CDF is not differentiable. At such points, the CDF has a different derivative when approached from the right as from the left. We are allowed to set the value of $f_X(x)$ arbitrarily to any nonnegative number at those few isolated points where the CDF is not differentiable. Note that this arbitrarily chosen value assigned to the pdf at these isolated points makes no difference whatsoever in any probability calculations, because the probability that this number occurs is zero. The probability that this number occurs is 0! In practice, we often choose either the derivative from the right or the derivative from the left as the value of the PDF at non-differentiable points of the CDF.

Example 3.6
Assume a continuous random variable has a CDF given by

$$F_X(x) = \begin{cases} 0 & x < -3, \\ \frac{1}{6}(x-3) & -3 \le x \le 3, \\ 0 & x > 3\ . \end{cases}$$

This density is illustrated in Figure 3.7. Compute $f_X(x)$, and define it for all $x \in \Re$.
Note that $F_X(x)$ is differentiable everywhere except at $x = \pm 3$. Then,

$$f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 0 & x < -3, \\ \frac{1}{6} & -3 < x < 3, \\ 0 & x > 3\ . \end{cases}$$

This is also shown in Fig. 3.8. To complete the definition, we select $f_X(3) = 0 = f_X(-3)$, which matches the slope of one of the two line segments that meet at 3 and -3.

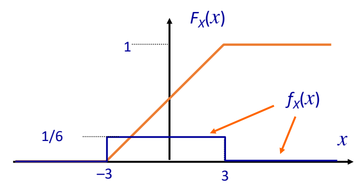

Figure 3.8: CDF and PDF for example 3.6.

As was the case for discrete random variables, we define the **expected value** of a continuous random variable X as

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

This is also known as the **mean** or **average.**. Similar to discrete random variables, this expected value can be viewed as the center of probability mass. If we repeat an experiment N times, add up all observed values of X, and divide by N to compute a sample average, the result should be pretty close to $\mathbb{E}[X]$. We sometimes use the notation $\mu_X = \mathbb{E}[X]$.

Note that, for the expectation to be defined, both of the integrals below must be finite.

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

This is not always the case, as shown in the next example.

Example 3.7
Let X be a continuous random variable with PDF given by: $f_X(x) = \begin{cases} \frac{2}{\pi(1+x^2)} & x \ge 0 \\ 0 & \text{otherwise.} \end{cases}$.

Note that this is a valid PDF, as it is nonnegative and properly normalized. It does decay to 0 slowly, in an inverse square law. For this random variable, its expected value does not exist:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{0}^{\infty} \frac{2x}{\pi(1+x^2)} d\ x = \ln(1+x^2)|_0^{\infty} = \infty\ .$$

This illustrates that some statistics of RVs may not be defined because the required expected values may not exist.

3.3.2 Variance

The **variance** measures how spread out a random variable is around its mean. For a continuous random variable X, it is defined using expectations in the same way as it was for discrete random variables:

$$\begin{aligned}\mathsf{Var}[X] &= \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 \\ &= \int_{-\infty}^{\infty} (x - \mu_X)^2 \, f_X(x) \, dx.\end{aligned}$$

We often refer to the variance of X as $\sigma_X^2 = \mathsf{Var}[X]$, where $\sigma_X \geq 0$ is the **standard deviation.**

3.3.3 Expected Value of a Function of a Random Variable

Let $g(\cdot)$ be a function mapping the range of a random variable X, R_X, into the real numbers $\Re$. Then, the variable $Y = g(X)$ is a random variable. We can compute the expected value of $Y = g(X)$ using the definition of the function and the PDF of X, as

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx \ .$$

Note that this expression is valid no matter whether the random variable Y is discrete, continuous or of other types. It avoids the need for computing the detailed PDF or PMF of Y, by performing the averaging in terms of the PDF of the random variable X.

Example 3.8
Let X be a continuous random variable with PDF $f_X(x) = \begin{cases} 0.5 & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$.

We compute μ_X as

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) d\,x = \int_{-1}^{1} 0.5 x d\,x = 0 \ .$$

The variance σ_X^2 is given by:

$$\mathsf{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-1}^{1} 0.5 x^2 d\,x = \frac{1}{3} \ .$$

Let $g(x) = |x|$, the absolute value function, and let $Y = g(X)$. Then,

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-1}^{1} 0.5|x| d\,x = \int_{-1}^{0} 0.5(-x) d\,x + \int_{0}^{1} 0.5(x) d\,x = 0.5 \ .$$

An important class of functions are the affine functions $g(x) = ax + b$. For these classes of functions, we establish the same relations that were established in 2.4. Let $Y = g(X) = aX + b$. Then, $\mathbb{E}[Y] = \mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b$. In addition, we can compute the variance as:

$$\mathsf{Var}[Y] = \mathbb{E}[(aX + b - a\mathbb{E}[x] - b)^2] = \mathbb{E}[(a(X - \mathbb{E}[X]))^2] = a^2 \mathsf{Var}[X] \ .$$

Thus, the variance of Y does not depend on the constant b, and is related to the variance of X as $\mathsf{Var}[Y] = a^2 \mathsf{Var}[X]$, as variance is a square statistic. Note that, in terms of standard deviation, $\sigma_Y = \sqrt{\mathsf{Var}[Y]} = |a|\sigma_X$, so that the standard deviation scales linearly with a.

Let $g(x) = a g_1(x) + b g_2(x)$, and let $Y = g(X)$. Then,

$$\mathbb{E}[Y] = \mathbb{E}[a g_1(X)] + \mathbb{E}[b g_2(X)] = a \int_{-\infty}^{\infty} g_1(x) d\,x + b \int_{-\infty}^{\infty} g_2(x) d\,x = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

emphasizing the fact that $\mathbb{E}[\cdot]$ is a linear operator.

3.3.4 Moments

Using the expectation operator, we define the following moments for continuous random variables, in exactly the same way they were defined for discrete random variables:

Definitions (same as for discrete random variables):

Mean of X	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$
Variance of X	$\mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2\, f_X(x)\, dx.$
n^{th} moment of X	$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n\, f_X(x)\, dx.$
n^{th} central moment of X	$\mathbb{E}[(X - \mathbb{E}[X])^n] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^n\, f_X(x)\, dx.$

Example 3.9
Assume a continuous random variable X which has a PDF given by

$$f_X(x) = \begin{cases} \frac{3}{2}x^2 & -1 \leq x \leq 1, \\ 0 & \text{elsewhere.} \end{cases}$$

This density is illustrated in Figure 3.9. Compute the mean, second moment, third moment and fourth central moment.
First, note the symmetry of $f_X(x)$ about zero. This means that, for any odd function where $f(x) = -f(-x)$, we have $\mathbb{E}[f(X)] = 0$. In particular, the first and third moments are expectations of odd functions $f(x) = x$ and $f(x) = x^3$, so we have $\mathbb{E}[X] = 0, \mathbb{E}[X^3] = 0$.
The second moment is

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) d\ x = \int_{-1}^{1} x^2 \frac{3}{2} x^2 d\ x = \frac{3}{10} x^5 |_{-1}^{1} = \frac{3}{5}\ .$$

Since the mean is zero, the fourth central moment is equal to the fourth moment:

$$\mathbb{E}[X^4] = \int_{-\infty}^{\infty} x^4 f_X(x) d\ x = \int_{-1}^{1} x^4 \frac{3}{2} x^2 d\ x = \frac{3}{14} x^7 |_{-1}^{1} = \frac{3}{7}\ .$$

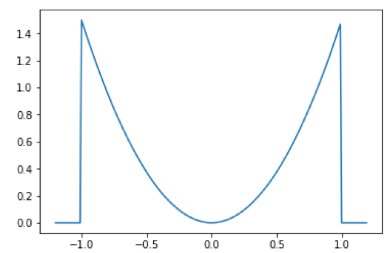

Figure 3.9: Figure for example 3.9.

3.4 Important Families of Continuous Random Variables

Although most experimental measurements are of limited precision, it is often easier to model their outcomes in terms of continuous-valued random variables because it facilitates the resulting analysis. Furthermore, the limiting form of many discrete-valued random variables result in continuous-valued random variables. Below, we describe some of the most useful continuous-valued random variables. Specifically, we overview the properties of the following families of continuous random variables:

- Uniform
- Exponential
- Gaussian (Normal)

These families of continuous RVs are used to model the outcomes of common experiments. Members of a given family differ only by the values of the few parameters of the family, which are easy to estimate from sample data. We also discuss a few other families of continuous random variables that are used less frequently in engineering applications.

3.4.1 Uniform(a, b) Random Variables

The simplest continuous random variable is the Uniform(a, b) random variable X, where X is equally likely to achieve any value in an interval of the real line, $[a, b]$. The probability density function of X is given by:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

The corresponding cumulative distribution function is given by

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

The PDF and CDF of uniform random variables are shown in Figure 3.10.

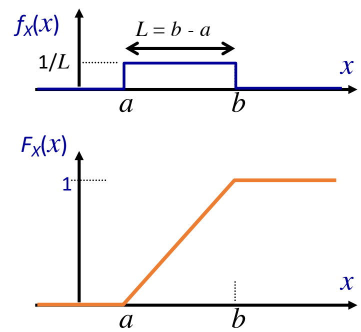

Figure 3.10: CDF and PDF for uniform RVs.

We use the notation $X \sim$ Uniform$([a, b])$ to denote a random variable with continuous uniform distribution on the interval a, b. Using the PDF, we compute the statistics of $X \sim$ Uniform$([a, b])$ as:

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) d\ x = \int_a^b \frac{x}{b-a} d\ x = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \quad \text{Mean} \\ \mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) d\ x = \int_a^b \frac{x^2}{b-a} d\ x = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3} \\ \mathsf{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12} \quad \text{Variance} \end{aligned}$$

Example 3.10
Consider a random wave of known amplitude A is oscillating at frequency ω_0 radians per second, but with unknown phase. We model the unknown phase as a random variable Θ, uniformly distributed on the interval $[-\pi, \pi]$, so that the time history of the wave is represented as

$$x(t) = Acos(\omega_0 t + \Theta).$$

From the properties of uniform random variables, we know the average phase $\mathbb{E}[\Theta] = 0$, and the variance of the phase is $\mathsf{Var}[\Theta] = \frac{(\pi - (-\pi))^2}{12} = \frac{\pi^2}{3}$.

The important statistics of uniform random variables are summarized below:

- PDF: $f_X(x) = \begin{cases} \dfrac{1}{b-a} & a\ lex \le b, \\ 0 & \text{otherwise.} \end{cases}$
- CDF: $F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x < b \\ 1 & b \le x\ . \end{cases}$
- Expected Value: $\mathbb{E}[X] = \dfrac{a+b}{2}$.

- Variance: $\mathsf{Var}[X] = \dfrac{(b-a)^2}{12}$.

- Interpretation: Equally likely to take any value between a and b.

3.4.2 Exponential(λ) Random Variables

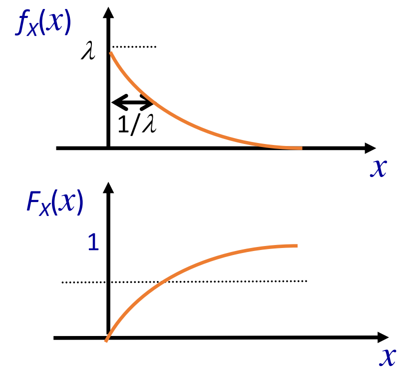

Figure 3.11: CDF and PDF for exponential RVs.

Exponential(λ) random variables arise in the modeling of the time between occurrence of events, such as the time between customer requests in service systems, the durations for call connections in phone systems, and the modeling of lifetimes of devices and systems. The exponential random variable X with parameter λ has a probability density function

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

The parameter λ is denoted as the rate of the exponential random variable, and it is typically measured as units per time. An exponential random variable only takes values in the non-negative real line. The corresponding CDF is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

The PDF and CDF of exponential random variables are shown in Figure 3.11.

The exponential random variable is similar to the discrete geometric random variable, in that it is the limit of the geometric random variable, as the difference between values of a geometric random variable gets small. For example, assume that an interval of length T seconds was subdivided into subintervals of length T/n, and assume that, for each subinterval, there is a Bernoulli trial with probability of success $p = \frac{\lambda T}{n}$, where λ is the average number of events per second, so λT is the average number of events per T seconds. Then, the number of subintervals until the occurrence of the next event is a geometric random variable M. Let X denote the time until the next successful event. Then, for any t which is a multiple of T/n,

$$\mathbb{P}[\{X > t\})] = \mathbb{P}[\{M > \frac{nt}{T}\}] = (1-p)^{nt/T} = \left((1 - \frac{\lambda T}{n})^n\right)^{t/T}$$

In the limit, we get

$$\lim_{n\to\infty} \mathbb{P}[\{X > t\}] = e^{-\lambda t}$$

which is $1 - F_X(t)$ for an exponential random variable X with rate λ.

We use the notation $X \sim$ exponential(λ) to denote a random variable X with exponential distribution, parameter λ. The important expectations of an exponential random variable $X \sim$ exponential(λ) are computed as:

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) d\,x = \int_0^{\infty} x\lambda e^{-\lambda x} d\,x \text{ (integrate by parts)} \\
&= -xe^{-\lambda x}|_0^{\infty} + \int_0^{\infty} \lambda e^{-\lambda x} d\,x \\
&= 0 - \frac{1}{\lambda} e^{-\lambda x}|_0^{\infty} = \frac{1}{\lambda} \\
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x f_X(x) d\,x = \int_0^{\infty} x^2 \lambda e^{-\lambda x} d\,x = -\int_0^{\infty} x^2 d\, e^{-\lambda x} \text{ (integrate by parts twice)} \\
&= x^2 e^{-\lambda x}|_0^{\infty} - 2\int_0^{\infty} xe^{-\lambda x} d\,x = \frac{2}{\lambda}\int_0^{\infty} xd\, e^{-\lambda x} \\
&= \frac{2}{\lambda} xe^{-\lambda x}|_0^{\infty} - \frac{2}{\lambda}\int_0^{\infty} e^{-\lambda x} d\,x = \frac{2}{\lambda^2} \\
\mathsf{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}
$$

Example 3.11
The duration of a service repair request for a broken appliance is modeled as an exponential random variable X with parameter $\lambda = 0.1$ repairs/minute. The repair person charges a fixed rate of \$5.00 for the first five minutes, then \$0.50 for each additional minute. Compute the expected time to repair the appliance, the variance of the repair time, and the expected cost of the repair.

Since X is an exponential random variable, the expected repair time and variance are computed as:

$$\mathbb{E}[X] = \frac{1}{\lambda} = 10\text{minutes. } \mathsf{Var}[X] = \frac{1}{\lambda^2} = 100\text{minutes}^2 .$$

The cost can be viewed as a function $g(X)$ defined by

$$g(x) = \begin{cases} 0 & x < 0, \\ 5 & 0 \le x \le 5, \\ 5 + 0.5(x-5) & x \ge 5 . \end{cases}$$

Then,

$$
\begin{aligned}
\mathbb{E}[g(X)] &= \int_0^{\infty} g(x) f_X(x) dx = \int_0^{\infty} 5 f_X(x) d\,x + \int_5^{\infty} 0.5(x-5) f_X(x) d\,x \\
&= 5 + \int_5^{\infty} 0.5(x-5) 0.1 e^{-0.1x} d\,x = 5\int_0^{\infty} 0.5(y) 0.1 e^{-0.1(y+5)} d\,y \text{ substitute } y = x-5 \\
&= 5 + 0.5e^{-0.5}\int_0^{\infty} 0.1 y e^{-0.1y} d\,y = 5 + 0.5e^{-0.5}\mathbb{E}[X] = 5 + 5e^{-0.5} \approx \$8.03 .
\end{aligned}
$$

Note that the expected cost $\mathbb{E}[g(X)]$ is not equal to $g(\mathbb{E}[X]) = \$7.50$. This is because $g(\cdot)$ is not an affine function.

The properties of exponential(λ) random variables are summarized below:

- PDF: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0, \\ 0 & x < 0. \end{cases}$
- CDF: $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 . \end{cases}$
- Expected Value: $\mathbb{E}[X] = \frac{1}{\lambda}$.
- Variance: $\mathsf{Var}[X] = \frac{1}{\lambda^2}$.

- Interpretation: Continuous waiting time. "Continuous version" of geometric random variables.
- Applications: Packet interarrival times, call durations, hard drive lifetimes.

3.4.3 Gaussian(μ, σ^2) Random Variables

Gaussian(μ, σ^2) random variables model many situations where the random event consists of the sum of a large number of small random variables. They are named after Karl Friedrich Gauss, who used this class of random variables to model errors in measurements for the least squares estimation of orbital parameters from telescope observations. Gaussian random variables are determined by two parameters: their mean μ and their variance σ^2.

The probability density function of a Gaussian random variable is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

. Its corresponding CDF is given by

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}} d\ u = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} d\ y\ .$$

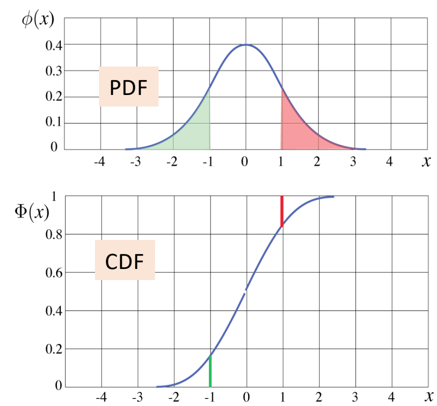

Figure 3.12: PDF and CDF for Gaussian RVs.

where the last equality follows by substituting $y = \frac{u-\mu}{\sigma}$. Note that the last integral corresponds to a Gaussian CDF win mean zero and variance 1. The PDF and CDF of Gaussian random variables are shown in Figure 3.12.

We refer to a Gaussian(0,1) random variable as a standard Gaussian random variable. Note that the CDF of any Gaussian(μ, σ^2) random variable can be computed in terms of the CDF of a standard Gaussian random variable. We formally define the CDF of a standard Gaussian as the function

$$\Phi(x) \equiv \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} d\ yn\ ,$$

and the standard normal complementary CDF as $Q(x) \equiv 1 - \Phi(x)$. Unfortunately, $\Phi(x)$ cannot be computed in closed form, but its values are tabulated in Appendix C.

Gaussian random variables are also known as Normal random variables because many sets of data gathered from a variety of physical phenomena seem to fit the Gaussian (or normal) distribution. In these sets of data, errors arise as the combination of many small effects; to develop the exact distribution of the sum of many random variables is unwieldy. Fortunately, the central limit theorem, which we study in Chapter 8, asserts that if many "small" random causes produce a net effect, then that effect can be approximately modeled as a normal or Gaussian random variable.

We often write $X \sim \mathcal{N}(\mu, \sigma^2)$, or use the phrase "$X$ is $\mathcal{N}(\mu, \sigma^2)$", to denote that X is a Gaussian(μ, σ^2) random variable with mean μ and variance σ^2. The statistics of a Gaussian random variable X are specified in its parameters:

$$\mathbb{E}[X] = \mu \qquad \mathsf{Var}[X] = \sigma^2\ .$$

We note that this notation varies across texts. Some texts will refer to a Gassian random variable as $\mathcal{N}(\mu, \sigma)$, using the standard deviation instead of the variance. We chose our notation because it generalizes to vectors in a natural way.

Normal distributions are used in many situations. In many classes, professors believe that the distribution of grades must be normally distributed with a given mean and variance. Thus, you see the phenomena that exams are graded "on the curve," where the actual grades are mapped nonlinearly into the Normal bell-shaped PDF, and letter grades are assigned based on the percentile of the grade using the standard Normal CDF $\Phi(x)$. Similarly, SAT and GRE actual scores are nonlinearly mapped so that the final scores correspond to a $\mathcal{N}(500, 10000)$ distribution.

Gaussian random variables have an interesting property: an affine transformation of a Gaussian random variable is also a Gaussian random variable. That is, if $X\ \mathcal{N}(\mu, \sigma^2)$ is Gaussian, then $Y = aX + b$ is also Gaussian for any real scalars a, b. We will show this later in this chapter. Furthermore, we know $\mathbb{E}[Y] = a\mathbb{E}[X] + b, \mathsf{Var}[Y] = a^2\mathsf{Var}[X]$, so

$$X\ N(\mu, \sigma^2) \rightarrow Y = aX + b\ \mathcal{N}(a\mu + b, a^2\sigma^2)$$

This important property is another reason why Gaussian variables are often used in engineering models.

The Gaussian PDF is symmetric about its mean. This implies that all odd central moments are zero. Using integration by parts, we can compute all even central moments as a multiple of the variance σ^2, as

$$\mathbb{E}[(X - \mathbb{E}[X])^{2n}] = (2n-1)(2n-3)\cdots(1)\sigma^2$$

. Thus, $\mathbb{E}[(X - \mathbb{E}[X])^4] = 3\sigma^2, \mathbb{E}[(X - \mathbb{E}[X])^6] = 15\sigma^2$.

To perform computations about probabilities of Gaussians, we use the standard normal CDF function $\Phi(\cdot)$. Appendix C includes the detailed tabulated standard normal CDF. We note the following properties which are useful for computation:

$$\Phi(-x) = 1 - \Phi(x) \qquad .\Phi(x) - \Phi(-x) = 2\Phi(x) - 1$$

$$Q(x) = \Phi(-x) = 1 - \Phi(x)$$

The way we use the standard tables for computation is for computing probabilities for a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Recall that

$$F_X(x) = \Phi(\frac{x-\mu}{\sigma})\ .$$

Note that the argument of the standard Gaussian function $\Phi(\cdot)$ is expressed as the difference between the value x and the mean of the random variable, expressed in units of standard deviations. That is, the statistic $z_x = \frac{x-\mu}{\sigma}$ used as the argument for Φ is the number of standard deviations away from the average. We illustrate this with the following example:

Example 3.12
Consider a Gaussian random variable $X \sim \mathcal{N}(1, 4)$. Determine the probability that X lies between -1 and 3.

From its definition, $\mathbb{P}[\{-1 < X \leq 3\}] = F_X(3) - F_X(1)$. The standard deviation of X is $\sigma_X = \sqrt{4} = 2$. Then,

$$\begin{aligned}
\mathbb{P}[\{-1 < X \leq 3\}] &= F_X(3) - F_X(-1)\\
z_3 &= \frac{3-1}{2} = 1;\ z_{-1} = \frac{-1-1}{2} = -1;\\
\mathbb{P}[\{-1 < X \leq 3\}] &= F_X(3) - F_X(-1) = \Phi(z_3) - \Phi(z_{-1}) = \Phi(1) - \Phi(-1)\\
&= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 = 2(0.8413)–1 = 0.6826
\end{aligned}$$

where the number for $\Phi(1)$ was obtained from the table in Appendix C.

Example 3.13
An underwater microphone is measuring the average acoustic pressure X to detect whether there is a submarine generating sounds in its neighborhood. If no submarine is present, the background acoustic pressure is modeled as a Gaussian, with $X \sim \mathcal{N}(2, 4)$. If the submarine is present, the measured acoustic pressure is modeled as $X \sim \mathcal{N}(3, 4)$.

The microphone uses a simple threshold $T \in (2, 3)$, and if the measured $X > T$, it declares that a submarine is present. A false alarm happens when there is no submarine present (so $X \sim \mathcal{N}(2, 4)$), yet $X > T$. What is the probability of a false alarm? Express the answer in terms of T and the standard complementary CDF $Q(\cdot)$.

When no submarine is present, $X \sim \mathcal{N}(2,4)$, so $\mu = \sigma = 2$. $P_F = \mathbb{P}[\{X > T\}] = 1 - F_X(T)$. Computing the z-statistic, $z_T = \frac{T-2}{2}$. Thus,

$$P_F = 1 - \Phi(z_T) = Q(z_T) = Q(\frac{T-2}{2}).$$

If the submarine is present, but $X < T$, the microphone will not declare a detection, and thus the detection will be missed. Express the probability of missed detection in terms of T using the complementary CDF $Q(\cdot)$.

When the submarine is present, $X \sim \mathcal{N}(3,4)$, so $\mu = 3, \sigma = 2$. Then, $P_{MD} = \mathbb{P}[\{X < T\} = F_X(T)$. The z-statistic is $z_T = \frac{T-3}{2}$, so

$$P_{MD} = F_X(T) = \Phi(z_T) = \Phi(\frac{T-3}{2}) = Q(\frac{3-T}{2})$$

A summary of the properties of a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is:

- PDF: $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
- CDF: $F_X(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\right)$ where $\Phi(z) = \displaystyle\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}\, d\, w$.
- $\Phi(z)$ is the standard normal CDF. $Q(z) = 1 - \Phi(z)$ is the standard normal complementary CDF.
- Expected Value: $\mathbb{E}[X] = \mu$.
- Variance: $\mathsf{Var}[X] = \sigma^2$.
- Interpretation: Sum (or average) of many small random effects.
- Applications: Noise modeling, linear systems, high-dimensional data.

3.4.4 Other families of continuous random variables

Below we quickly overview other classes of continuous random variables that are used less frequently in engineering. This section is primarily for reference, and won't be used much in the rest of this course.

Gamma and Erlang random variables Gamma random variable appear in may applications. For example, it is often used to model the time to service customers in queuing systems, the lifetime of devices in reliability studies, and the defect clustering behavior in VLSI chips. The probability density function of a gamma random variable X has two parameters $\rho > 0, \lambda > 0$, and is given by

$$f_X = \frac{\alpha(\alpha x)^{\rho-1} e^{-\lambda x}}{\Gamma(\rho)}$$

where $\Gamma(z) = \displaystyle\int_0^{\infty} x^{z-1} e^{-x} d\, x$. Note that, for z a positive integer, $\Gamma(z) = (z-1)!$. Other notable values are $\Gamma(0.5) = \sqrt{\pi}$.

The versatility of the gamma distribution is that, by properly choosing the two parameters, it can take a variety of shapes, which can be used to fit specific distributions. For instance, when $\rho = 1$, we obtain the exponential random variable. By letting $\rho = m$, where m is a positive integer, we obtain the m-stage Erlang distribution, which is the distribution of the sum of m independent and identical exponential random variables, each with rate λ.

The CDF of general Gamma distributions can only be expressed in terms of special functions and is seldom used for computations. The important expectations of a gamma random variable X with parameters ρ, λ are given by:

$$\mathbb{E}[X] = \frac{\rho}{\lambda}$$
$$\mathsf{Var}[X] = \frac{\rho}{\lambda^2}$$

Rayleigh random variables Rayleigh random variables are often used to model the random magnitude of a vector. As such the variables must be positive. The PDF of a Rayleigh random variable X with parameter α is given by: $f_X(x) = \dfrac{x}{\alpha^2} e^{-x^2/2\alpha^2}$.

The CDF of a Rayleigh random variable X with parameter α is $F_X(x) = 1 - e^{-\frac{x^2}{2\alpha^2}}$. Some of its important statistics are

$$\mathbb{E}[X] = \alpha\sqrt{\frac{\pi}{2}}$$
$$\mathsf{Var}[X] = \frac{4-\pi}{2}\alpha^2$$

Laplacian random variable: The Laplacian random variable models a two-sided exponential distribution with parameter λ. The probability density function of a Laplacian random variable X is given by

$$f_X(x) = \frac{\lambda}{2} e^{-\lambda|x|} \; .$$

Its CDF is given by

$$F_X(x) = \begin{cases} \frac{1}{2} e^{\lambda x} & x < 0, \\ 1 - \frac{1}{2} e^{-\lambda x} & x \geq 0 \; . \end{cases}$$

with expectations:

$$\mathbb{E}[X] = 0$$
$$\mathsf{Var}[X] = \frac{2}{\lambda^2}$$

Cauchy random variable: The Cauchy random variable is often used as an example to illustrate distributions which do not decay fast enough as $x \to \infty$, so that no moments exist. We call those *heavy-tailed* distributions. The probability density function of a Cauchy random variable with parameter β is given by $f_X = \dfrac{\beta/\pi}{\beta^2 + x^2}$.

The CDF of a Cauchy random variable X with parameter β is $F_X(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\tan^{-1}(\dfrac{x}{\beta})$.

Due to its symmetry, the mean is often taken to be zero, though the formal expected value of the density does not have a unique value. It is easy to verify that the variance of this distribution does not exist either.

In Table 3.2 we summarize the characteristics of important random variables, where the more general (shifted) forms of the Laplacian and Cauchy distributions are given.

Example 3.14
Consider the following quick questions regarding continuous random variables:

Discrete-Valued X						
Name	Range R_X	Parameters	PMF $P_X(x)$	CDF $F_X(x)$	Mean	Variance z)
Bernoulli	$\{0,1\}$	$0 \leq p \leq 1$	$P(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \end{cases}$	$\begin{cases} 0 & x<0, \\ (1-p) & x \in [0,1) \\ 1 & x>1. \end{cases}$	p	$p(1-p)$
Uniform	$\{k, k+1, \ldots, k+n\}$	$n>0, k$ integer	$\frac{1}{n+1}$	$\frac{\lfloor x \rfloor - k + +1}{n+1}$	$k+\frac{n}{2}$	$\frac{(n+1)^2-1}{12}$
Binomial	$\{0, \ldots, n\}$	$0 \leq p \leq 1$	$\binom{n}{x} p^x (1-p)^{(n-x)}$	$\sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{(n-k)}$	np	$np(1-p)$
Geometric	$\{1, \ldots\}$	$0<p<1$	$(1-p)^{x-1}p$	$(1-(1-p)^{\lfloor x \rfloor})$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$\{0, 1, \ldots\}$	$0<\lambda$	$\frac{\lambda^x e^{-\lambda}}{x!}$	$\sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ

Continuous-Valued X						
Name	Range	Parameters	PDF $f_X(x)$	CMF $F_X(x)$	Mean	Variance
Uniform	$[a,b]$	$a<b$	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Gaussian	$[-\infty, \infty]$	μ, σ^2	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$	$1-Q((x-\mu)/\sigma)$	μ	σ^2
Exponential	$[0, \infty]$	$\lambda>0$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Erlang	$[0, \infty]$	$\lambda>0,\ n>0$	$\frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$	$1-e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
Gamma	$[0, \infty]$	$\lambda, \rho > 0$	$\frac{\lambda(\lambda x)^{(\rho-1)} e^{-\lambda x}}{\Gamma(\rho)}$	N/A	$\frac{\rho}{\lambda}$	$\frac{\rho}{\lambda^2}$
Rayleigh	$[0, \infty]$	α^2	$\frac{x}{\alpha^2} e^{-x^2/2\alpha^2}$	$1-e^{-x^2/2\alpha^2}$	$\alpha\sqrt{\frac{\pi}{2}}$	$(2-\frac{\pi}{2})\alpha^2$
Laplacian	$[-\infty, \infty]$	$\lambda>0, \mu$	$\frac{\lambda}{2} e^{-\lambda \lvert x-\mu \rvert}$	$\begin{cases} \frac{1}{2} e^{\lambda(x-\mu)} & x<\mu \\ 1-\frac{1}{2} e^{-\lambda(x-\mu)} & x \geq \mu \end{cases}$	μ	$\frac{2}{\lambda^2}$
Cauchy	$[-\infty, \infty]$	$\beta>0, \alpha$	$\frac{\beta/\pi}{\beta^2+(x-\alpha)^2}$	$\frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(\frac{x-\alpha}{\beta}\right)$	Undef	Undef

Table 3.2: Important random variables. (N/A under the PDF column indicates that there is no simplified form.)

1. if $X \sim$ Uniform($[0,1]$), and $Y = -2X + 1$, compute $\mathbb{E}[Y]$ and Var$[Y]$.
   Answer: $\mathbb{E}[X] = \frac{0+1}{2} = 0.5; \mathbb{E}[Y] = -2\mathbb{E}[X] + 1 = 0; \text{Var}[X] = \frac{1}{12}; \text{Var}[Y] = (-2)^2\text{Var}[X] = \frac{1}{3}$.
2. If $X \sim$ Uniform($[a,b]$), and $\mathbb{E}[X] = 2, \text{Var}[X] = 4$, what are a, b?
   Answer: $\mathbb{E}[X] = 2 = \frac{a+b}{2}; \text{Var}[X] = 4 = \frac{(b-a)^2}{12}$ so $b - a = 4\sqrt{3}$. Thus, $b = 2 + 2\sqrt{3}, a = 2 - 2\sqrt{3}$.
3. If $X \sim \mathcal{N}(0, 0.5)$ and $Y = -2X + 1$, what is the probability that $Y > 5$, in terms of the standard Gaussian CDF $\Phi(x)$?
   Answer: $\mathbb{E}[Y] = 1, \text{Var}[Y] = (-2)^2(0.5) = 2$. Thus, $\mathbb{P}[\{Y > 5\}] = 1 - F_Y(5)$. Since Y has mean 1, standard deviation $\sqrt{2}$, the z statistic $z_5 = \frac{5-1}{\sqrt{2}} = 2\sqrt{2}$. Hence, $\mathbb{P}[\{Y > 5\}] = 1 - \Phi(2\sqrt{2}) = Q(2\sqrt{2})$.
4. If X is Gaussian with $\mathbb{E}[X] = 1, \mathbb{E}[X^2] = 5$, what is the variance of X?
   Answer: $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 4$.

3.5 Conditional Probability for Continuous Random Variables

Consider a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. For any events $A, B \in \mathcal{E}$ such that $\mathbb{P}[B] > 0$, we define the conditional probability of A given B as:

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B - A] + \mathbb{P}[B \cap A]} \ .$$

Let X be a random variable defined on $(\Omega, \mathcal{E}, \mathbb{P})$. Then, $\{\omega \in \Omega : X(\omega) \leq a\} \equiv \{X \leq a\}$ defines an event in $\mathcal{E}$. Using this event, we can define the conditional cumulative distribution function $F_{X|B}(a)$ as follows:

$$F_{X|B}(a) = \frac{\mathbb{P}[\{\omega \in \Omega : X(\omega) \leq a\} \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[\{X \leq a\} \cap B]}{\mathbb{P}[B]} \ .$$

Note that this definition is valid for all random variables, not just discrete or continuous ones. For discrete random variables, we defined the conditional probability mass function in **??**, as $P_{X|B}(a)$ by exploiting the fact that R_X was discrete:

$$P_{X|B}(a) = \frac{\mathbb{P}[\{X = a\} \cap B]}{\mathbb{P}[B]} \ .$$

For the special case that $B \subset X$, so that the event is $\{\omega \in \Omega : X(\omega) \in B\}$, this simplified to

$$P_{X|B}(a) = \begin{cases} \frac{P(a)}{\mathbb{P}_X[B]} & a \in B_1, \\ 0 & \text{otherwise.} \end{cases}$$

We referred to this operation as restrict/rescale: restrict the probability mass functions to $a \in B$, and rescale so that $\sum_{x \in B} P_{X|B}(x) = 1$.

Suppose X is a continuous random variable, so it has a probability density function $f_X(x)$ defined almost everywhere. We can compute the conditional CDF $F_{X|B}(a)$ as indicated above. Then, we define the conditional probability density function $f_{X|B}(a)$ as the derivative of the conditional CDF:

$$f_{X|B}(A) = \frac{d}{da} F_{X|B}(a) = \frac{\frac{d}{da}\mathbb{P}[\{X \leq a\} \cap B]}{\mathbb{P}[B]} = \frac{\lim_{\epsilon \to 0} \frac{\mathbb{P}[\{X \leq a+\epsilon\} \cap B] - \mathbb{P}[\{X \leq a\} \cap B]}{\epsilon}}{\mathbb{P}[B]}$$

It should be clear that, if X has a CDF that is differentiable almost everywhere, the conditional CDF will also be differentiable almost everywhere, so the conditional PDF will exist as defined above.

We can simplify this when the conditioning event is $\{X \in B \subset \Re\}$. In this case, $\mathbb{P}_X[B] = \int_{x \in B} f_X(x) d\, x = \mathbb{P}[\{X \in B\}]$. For this case, we have the following:

$$\mathbb{P}[\{X \leq a + \epsilon\} \cap \{X \in B\}] - \mathbb{P}[\{X \leq a\} \cap \{X \in B\}] = \int_{x \in (-\infty, a+\epsilon] \cap B} f_X(x) d\, x - \int_{x \in (-\infty, a] \cap B} f_X(x) d\, x \ .$$

Thus, taking limits and using the fundamental theorem of calculus,

$$\lim_{\epsilon \to 0} \frac{\mathbb{P}[\{X \leq a+\epsilon\} \cap \{X \in B\}] - \mathbb{P}[\{X \leq a\} \cap \{X \in B\}]}{\epsilon} = \begin{cases} 0 & \text{if } a \notin B, \\ \lim_{\epsilon \to 0} \frac{\mathbb{P}[\{X \leq a+\epsilon\}] - \mathbb{P}[\{X \leq a\}]}{\epsilon} & \text{if } a \in B. \end{cases}$$

Thus, when the conditioning event is $\{X \in B\}$, we have $f_{X|B}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbb{P}_X[B]} & x \in B \\ 0 & \text{otherwise.} \end{cases}$

Example 3.15
Let $X \sim$ Exponential(2) be an exponential random variable with rate 2. Consider the event B generated by $\{X > 1\}$. Then, $\mathbb{P}_X[B] = 1 - F_X(1) = e^{-2}$, and the conditional PDF of X given B is

$$f_{X|B}(x) = \begin{cases} \frac{2e^{-2x}}{e^{-2}} = 2e^{-2(x-1)} & x \geq 1, \\ 0 & x < 1 \ . \end{cases}$$

Note that the conditional $f_{X|B}(x)$ is just the original $f_X(x)$ shifted to start at $x = 1$! This is the memoryless property for exponential random variables that we showed earlier for geometric random variables. If we define the time to go as $Y = X - 1$, then $f_{Y|B}(a) = f_X(a)$. Thus, if you have waited for one hour for an arrival, the time you have left to wait has the same distribution as the original arrival time.

With the conditional PDF, we can define conditional statistics: The conditional expected value of X is

$$\mathbb{E}[X|B] = \int_{-\infty}^{\infty} x f_{X|B}(x) d\ x \ .$$

The conditional expected value of a function $g(X)$ is

$$\mathbb{E}[g(X)|B] = \int_{-\infty}^{\infty} g(x) f_X(x) d\ x \ .$$

With these equations, we can now compute the conditional variance of X given observation of event B as $\mathsf{Var}[X|B] = \mathbb{E}[X^2|B] - (\mathbb{E}[X|B])^2$.

3.6 Functions of a Continuous Random Variable

Assume we have a random variable X defined on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. Any measurable function $g : \Re \to \Re$ can be used to define a derived random variable $Y = g(X)$ on the same probability space. However, even if X is a continuous random variable, it is unclear as to whether the resulting random variable Y will be continuous, or discrete, or perhaps a mixed random variable, a type that we have not discussed yet. Even if X is a continuous random variable and $g(\cdot)$ is a continuous function, the resulting random variable Y is not guaranteed to be continuous.

We have described previously how to compute statistics of Y, such as $\mathbb{E}[Y]$ or $\mathbb{E}[Y^2]$, without having to compute the resulting PMF or PDF of Y; e.g. $\mathbb{E}[Y] = \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x) d\ x$. What we are after in this section is computing the full PMF or PDF of Y, whenever it is appropriate to do so.

Example 3.16
Let $X \sim$ Uniform($[-1, 1]$). Define $g(x) = \begin{cases} 0 & x < 0, \\ x & x \geq 0 \ . \end{cases}$. Note that $g(x)$ is continuous, and X is a continuous random variable, but $Y = g(X)$ has the property that $\mathbb{P}[\{Y = 0\}] = 0.5$, so that there is mass at the point $Y = 0$. Thus, Y is not a continuous random variable, because the CDF of Y is not continuous.

3.6.1 Transforming Continuous to Discrete

One case in which we can handle the transformation $Y = g(X)$ is whenever the function $g(\cdot)$ is piecewise constant. In this case, $Y = g(X)$ will be a discrete random variable, with range R_Y written as the list of discrete values that $g(x)$ can take. In this case, we can determine the PMF of Y as follows:

For each value $y \in R_Y$, determine the set of values of x such that $g(x) = y$. Formally, find $A_y = \{x : g(x) = y\}$ for each $y \in R_Y$. Then, compute $P_Y(y) = \int_{x \in A_y} f_X(x)d\ x$. The resulting random variable is discrete, and its statistics can be obtained for the PMF function computed above.

3.6.2 Transforming Continuous to Continuous

If X is a continuous random variable, the function $g(x)$ is continuous, differentiable almost everywhere, and its derivative $g'(x) = \frac{d}{dx}g(x)$ is not zero on any interval (but can be zero at specific values), then $Y = g(X)$ is a continuous random variable. Under these conditions, the set of values x such that $g(x) = y$ is discrete, and has probability zero. Thus, probability mass does not accumulate at any value of y.

For these cases, the PDF of Y can be determined from the PDF of X and knowledge of the function $g(\cdot)$ and its derivative. Given the range R_X, compute the range $R_Y = g(R_X)$. For each $y \in R_Y$, determine the set of all values of $x \in R_X$ such that $g(x) \leq y$. Formally, find $B_y = \{x \in R_X : g(x) \leq y\}$ for each $y \in R_Y$. Then, the CDF of Y is determined as

$$F_Y(y) = \mathbb{P}[\{Y \leq y\} = \int_{B_y} f_X(x)d\ x\ .$$

Once the CDF of Y is found, the PDF is obtained as the derivative $f_Y(y) = \frac{d}{dy}F_Y(y)$.

There is a special case of functions $g(\cdot)$ for which the computation of the PDF of Y is simpler, and can be done avoiding the need to compute the CDF of Y first. That is the case where $g(\cdot)$ is **strictly monotonic**: either strictly increasing ($g(x) > g(y)$ if $x > y$) or strictly decreasing ($g(x) < g(y)$ if $x > y$). In this case, the function $g(x)$ has an inverse function $h(y) = g^{-1}(y)$, and the PDF of $Y = g(X)$ is $f_Y(y) = f_X\big(h(y)\big)\left|\frac{d}{dy}h(y)\right|$.

A special case of monotone functions $g(\cdot)$ are affine functions $g(x) = ax + b$ where the slope a is non-zero. In this case, $h(y) = \frac{1}{a}(y - b)$, $\frac{d}{dy}h(y) = \frac{1}{a}$ and

$$f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)\ .$$

We illustrate this with several important examples:

Example 3.17
Let X be a Gaussian random variable such that $X \sim \mathcal{N}(\mu, \sigma^2)$. Let $g(x) = ax + b$, with $a \neq 0$. Then, $Y = g(X)$ is an affine transformation of a Gaussian random variable. By the above formula,

$$f_Y(y) = \frac{1}{|a|}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\frac{y-b}{a}-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi a^2\sigma^2}}e^{-\frac{(y-b-\mu a)^2}{2a^2\sigma^2}} \sim \mathcal{N}(a\mu + b, a^2\sigma^2)\ .$$

This proves the important property that affine transformations of Gaussian random variables are Gaussian random variables.

Example 3.18
Let X be a uniform random variable, with $X \sim$ Uniform((0, 1]). Let $g(x) = -\frac{1}{\lambda}\ln(x)$ which is a monotone, strictly decreasing function with inverse $h(y) = e^{-\lambda y}$. Let $Y = g(X)$; then, $R_Y = (0, \infty)$; then,

$$f_Y(y) = f_X\big(h(y)\big)\left|\frac{d}{dy}h(y)\right| = |\frac{d}{dy}h(y)| = \lambda e^{-\lambda y}, y > 0\ .$$

This shows $Y \sim$ Exponential(λ), an exponential random variable.

Example 3.19
Consider a function $q(y)$ that is continuous, monotone non-decreasing, differentiable almost everywhere with values in $[0,1]$ such that $\lim_{y\to-\infty} q(y) = 0, \lim_{y\to\infty} q(y) = 1$. Assume that $q(y)$ is strictly monotone increasing over its range $R_Y = \{y \in \Re : 0 < q(y) < 1\}$. Let X be a uniform random variable on $[0,1]$. We want to find a transformation $Y = g(X)$ such that the derived random variable Y has CDF $F_Y(y) = q(y)$.

Let $r(y) = q^{-1}(y)$ be the inverse of q, so that $r : (0,1) \to R_X$, and define $Y = r(X)$. Then,

$$F_Y(y) = \mathbb{P}[\{Y \le y\}] = \mathbb{P}[\{X \le q(y)\}] = F_X(q(y)) = q(y)$$

Hence, we can transform uniform random variables on [0,1] to random variables Y with CDF $q(Y)$ as long as $q(Y)$ is strictly increasing over its effective range.

Example 3.20
Let $X \sim \mathcal{N}(0,1)$, and let $Y = X^2$. Note that $g(\cdot)$ is continuously differentiable, but not monotone. For any value $y \in [0,\infty)$, let $B_y = \{x \in \Re : x^2 \le y\} = .\{x \in \Re : -\sqrt{y}x \le \sqrt{y}\}$. Then,

$$F_Y(y) = \mathbb{P}_X[B_y] = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \ .$$

Hence, its PDF for $y > 0$ is

$$f_Y(y) = \frac{d}{dy}\left(\Phi(\sqrt{y}) - \Phi(-\sqrt{y})\right) = \frac{1}{2\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}} + \frac{1}{2\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}} = \frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}} \ .$$

3.7 Mixed Random Variables

There are many random variables that are neither continuous nor discrete. For instance, consider the random variable X with CDF given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.5x & 0 \le x < 1 \\ 1 & x \ge 1 \ . \end{cases}$$

This CDF is not continuous, as it has a jump at $x = 0$. However, the range of X is $R_X = [0,1]$, an uncountable space. Random variables with a CDF that has a discrete set of discontinuities, but is almost surely differentiable elsewhere are mixtures of discrete and continuous random variables. We refer to such random variables as **mixed random variables.**

The difficulty with mixed random variables X is that we cannot compute either a probability mass function or a probability density function from the CDF $F_X(x)$. Hence, we don't have the basic information needed for computing statistics, or expectations of functions of X.

We will overcome this difficulty by defining a generalized version of a CDF using generalized derivatives of the CDF. Specifically, at points where the CDF has discontinuities, we represent the derivative using an impuse $\delta(\cdot)$ function. In engineering, the impulse function is defined by the following properties:

$$\delta(a) = 0 \text{if } a \ne 0$$

$$\int_b^c \delta(a)\, d\ a = \begin{cases} 0 & \text{if } b \le c < 0 \\ 1 & \text{if } b \le 0 \le c \\ 0 & \text{if } 0 < b \le c \ . \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(a-s) g(\omega)\, d\ s = g(a) \quad \text{if } g \text{ is continuous at } a.$$

Using this concept, we define the PDF of a mixed random variable X as: $f_X(x) = \frac{d}{dx}F_X(x)$ where we use impulse functions to represent derivatives at points where the CDF is discontinuous. Note we can use this do define a PDF for discrete random variables also. For example,

$$f_X(x) = 0.5\delta(x+1) + 0.5\delta(x-1)$$

is the density of a random variable taking on the values $\{-1, 1\}$ each with equal probability.

The most important property of impulse functions is that we can integrate them. We use PDFs to compute probabilities of events by integrating the PDF over the range of values in the event. Thus, for the above variable X, we can compute the second moment directly, as

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)d\ x = \int_{-\infty}^{\infty} x^2(0.5\delta(x+1) + 0.5\delta(x-1))d\ x = 0.5(-1)^2 + 0.5(1)^2 = 1\ .$$

Similarly, assume that the CDF of a mixed random variable X is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.5x & 0 \le x < 1 \\ 1 & x \ge 1\ . \end{cases}$$

The PDF can be computed as

$$f_X(x) = 0.5\delta(x) + 0.5I_{\{x\in(0,1(\}}$$

where the indicator function $I_{\{x\in(0,1(\}} = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}.$

Note that we still maintain the fundamental relationships between CDF and PDF:

$$F_X(a) = \int_{-\infty}^{a} f_X(\omega)ds.$$

Furthermore, for random variables $Y = g(X)$, we still have

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)d\ x\ .$$

whenever the integrals are finite and well-defined.

Example 3.21

A service station has two servers that it can use to handle services: a robot that always completes its service in 10 seconds, and a human that completes its service in a random time, distributed uniformly between 5 and 15 seconds. When you request service, you will be assigned the robot with probability 0.6, and the human with probability 0.4. Let X denote the random variable representing the time at which your service request will be completed. Note that X can take values between 5 and 15 seconds, a continuous interval.

What is the CDF of X? Let's compute this using the Law of Total Probability. Let B_1 be the set of all outcomes where the robot performs your service, and B_2 be the set of all outcomes where the human performs the service. Then, B_1, B_2 is a partition of all the possible outcomes. Using the Law of Total Probability,

$$F_X(x) = \mathbb{P}[\{X \le x\}] = \mathbb{P}[\{X \le x\}|B_1]\mathbb{P}[B_1] + \mathbb{P}[\{X \le x\}|B_2]\mathbb{P}[B_2]$$

From the information in the problem, $\mathbb{P}[B_1] = 0.6, \mathbb{P}[B_2] = 0.4$. We are also given

$$\mathbb{P}[\{X \le x\}|B_1] = \begin{cases} 0 & x < 10 \\ 1 & x \ge 10\ . \end{cases} ;\quad \mathbb{P}[\{X \le x\}|B_2] = \begin{cases} 0 & x < 5 \\ \frac{x-5}{10} & 5 \ge x < 10 \\ 1 & x \ge 10\ . \end{cases}$$

The CDF of X is thus obtained by direct substitution into the formula above. It is clearly the CDF of a mixed random variable, as it has a discontinuity at $x = 10$.

We compute the PDF of X as the derivative of this CDF, as

$$f_X(x) = 0.6\delta(x-10) + 0.04 I_{\{x\in[5,15]\}} \ .$$

We can now compute the expected value of X and its variance as:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) d\ x = 0.6 \int_{-\infty}^{\infty} x\delta(x-10)dx + 0.04 \int_5^{15} x dx = 6 + 4 = 10 \\
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) d\ x = 0.6 \int_{-\infty}^{\infty} x^2\delta(x-10)dx + 0.04 \int_5^{15} x^2 dx \\
&= 60 + 0.04 \frac{15^3 - 5^3}{3} = 60 + 130/3 \\
\mathsf{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 60 + 130/3 - 100 = \frac{10}{3}
\end{aligned}$$