

Chapter 5

Continuous Random Variables

In the previous chapter, we studied discrete random variables. This chapter is dedicated to another class of random variables, known as *continuous* random variables. A continuous random variable takes its values within entire intervals or unions of intervals, and its distribution can be identified by its probability *density* at each point within those intervals. Most aspects of dealing with continuous random variables are analogous to those regarding discrete random variables, once we replace summation with integration.

5.1 Why probability mass function is not sufficient

Recall that the distribution of a discrete random variable X can be described either by its pmf

$$p_X(x) := \mathbb{P}(X = x)$$

or by its cdf

$$F_X(x) := \mathbb{P}(X \leq x) .$$

While the cdf can be used to describe the distribution of *any* type of random variable, the pmf is not well-suited for non-discrete random variables.

Example 5.1.1 (A uniform RV). Consider the experiment of picking a number from the interval $[0, 1]$ at random (see Example 3.2.1). Let X stand for the picked number. This is a random variable¹ which is distributed uniformly over the interval $[0, 1]$. The idea that the distribution of X is uniform over $[0, 1]$ can be expressed by saying that

- (i) The possible values of X are the numbers in $[0, 1]$,
- (ii) For each $0 \leq a \leq b \leq 1$, $\mathbb{P}(X \in [a, b]) = \frac{b-a}{1-0} = b - a$.

Observe that, for every $x \in [0, 1]$,

$$\mathbb{P}(X = x) = \mathbb{P}(X \in [x, x]) = x - x = 0 .$$

which means that the probability mass function of X is everywhere zero, and hence bears hardly any information about the distribution of X . The cumulative distribution function of X , on the other hand, contains all the information about the distribution of X , as we shall see below.

Ⓚ What is the cdf of X ?

Ⓜ

$$F_X(x) := \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } 1 < x, \end{cases}$$

(see Figure 5.1a). If $x < 0$, then the event $X \leq x$ never happens, so $\mathbb{P}(X \leq x) = 0$. If $0 \leq x \leq 1$, then $\mathbb{P}(X \leq x) = \mathbb{P}(X \in [0, x]) = x$. If $x > 1$, the event $X \leq x$ always happens, thus $\mathbb{P}(X \leq x) = 1$.

Observe that in this example, the cdf is a continuous function.

Ⓚ How can we recover $\mathbb{P}(X \in [a, b])$ from the cdf of X ?

¹More specifically, $X(\omega) := \omega$ for each outcome $\omega \in [0, 1]$.

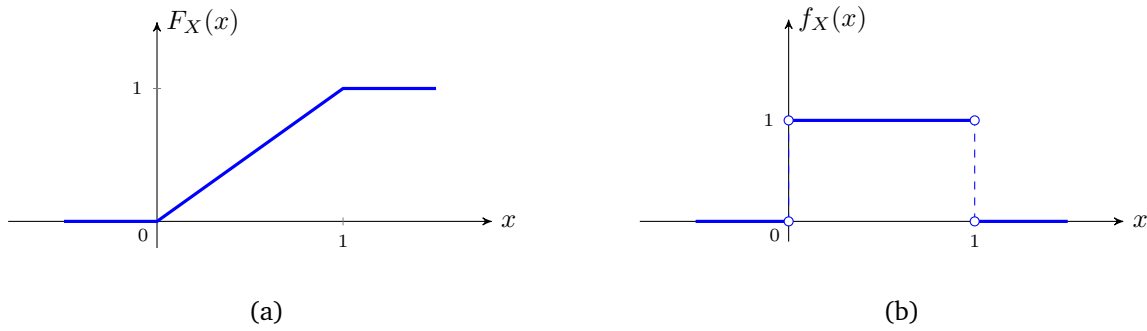


Figure 5.1: The distribution of a uniform RV over $[0, 1]$. (a) cdf (b) pdf

A Observe that if a, b are two numbers satisfying $a \leq b$, we have

$$F_X(b) - F_X(a) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = \mathbb{P}(a < X \leq b) = \mathbb{P}(X \in (a, b]) .$$

This is not quite the same as $\mathbb{P}(X \in [a, b])$, but it is a good start.

We can use the following trick to find $\mathbb{P}(X \in [a, b])$ (see Figure 5.2): Take a' smaller than a but very close to it. The chance that $a' < X \leq b$ is small, thus we can approximate $\mathbb{P}(a \leq X \leq b)$ with $\mathbb{P}(a' < X \leq b)$. The closer a' is to a , the better the approximation. In particular, in the limit we get

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(a \leq X \leq b) = \lim_{a' \uparrow a} \mathbb{P}(a' < X \leq b) = F_X(b) - \lim_{a' \uparrow a} F_X(a') .$$

Now, since in this example $F_X(x)$ is a continuous function, we have $\lim_{a' \uparrow a} F_X(a') = F_X(a)$. Hence,

$$\mathbb{P}(X \in [a, b]) = F(b) - F(a) ,$$

which happens to be the same as $\mathbb{P}(X \in (a, b])$. In particular, depending on the values of a, b , we get

$$\mathbb{P}(X \in [a, b]) = \begin{cases} 0 & \text{if } a \leq b < 0, \\ b & \text{if } a < 0 \leq b \leq 1, \\ 1 & \text{if } a < 0 < 1 < b, \\ b - a & \text{if } 0 \leq a \leq b \leq 1, \\ 1 - a & \text{if } 0 \leq a \leq 1 < b, \\ 0 & \text{if } 1 < a \leq b, \end{cases}$$

which completely characterizes the distribution of X as described earlier. ○



Figure 5.2: The probability that $X \in [a, b]$ can be approximated by the probability that $X \in (a', b]$, where a' is smaller than a but very close to it.

The above trick for deriving the probability that a random variable falls within a closed interval works in general.



Fact. Let X be a random variable (of any type), and let $F_X(x)$ be its cdf. Then, for every closed interval $[a, b]$, we have

$$\mathbb{P}(X \in [a, b]) = F_X(b) - \lim_{a' \uparrow a} F_X(a') .$$

Note, however, that $\lim_{a' \uparrow a} F_X(a') = F_X(a)$ only if F_X is continuous at a , thus $\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a)$ only if F_X is continuous at a , which is the case, precisely when $\mathbb{P}(X = a) = 0$. For instance, if X is a Bernoulli random variable with parameter p (Example 4.2.4), then

$$\mathbb{P}(X \in (0, 1/2]) = F_X(1/2) - F_X(0) = (1 - p) - (1 - p) = 0 ,$$

whereas

$$\mathbb{P}(X \in [0, 1/2]) = F_X(1/2) - \lim_{x \uparrow 0} F_X(x) = (1 - p) - 0 = 1 - p .$$

5.2 Distribution of a random variable: part II

Example 5.2.1 (A uniform RV). As in Example 5.1.1, let X be a random variable uniformly distributed over the interval $[0, 1]$. Suppose we repeat the underlying experiment many many times.

Ⓚ How will the histogram of the values of X in these repeated experiments look like?

Ⓐ Figure 5.3 shows a histogram of the values of X in $n = 100\,000$ repeated experiments.²

To understand the shape of the histogram, suppose we repeat the experiment n times, where n is very large. Let $B = [x, x + \varepsilon)$ be one of the bins of the histogram, which has width ε . Then,

$$\langle \text{area of the rectangle above } B \rangle = \frac{N_n(X \in B)}{n} \approx \mathbb{P}(X \in B),$$

where as usual, $N_n(X \in B)$ denotes the number of times in which $X \in B$. If B is inside $[0, 1]$, then $\mathbb{P}(X \in B) = \varepsilon$. Since the width of the rectangle above B is ε , its height will approximately be 1. If on the other hand, B is outside $[0, 1]$, then $\mathbb{P}(X \in B) = 0$, hence in none of the experiments will X be in B , so the height of the rectangle above B will be 0.

In conclusion, as the number of repeated experiments becomes larger and larger, the histogram will look more and more like Figure 5.1b.

The “idealized histogram” depicted in Figure 5.1b is called the *probability density function* of X . ○

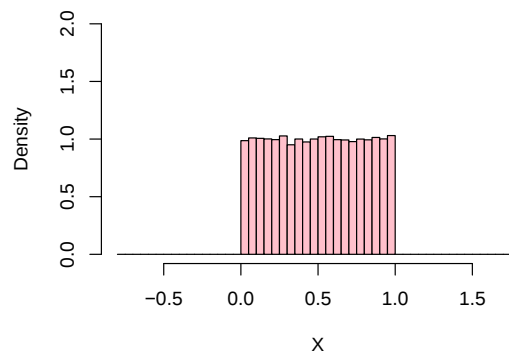


Figure 5.3: Histogram of the values of a RV uniformly distributed over $[0, 1]$ in 100 000 repeated experiments

Continuous random variables are those random variables whose distribution can be described by such “idealized histograms”.



Terminology. A random variable X is said to have *probability density function* (often abbreviated as *pdf*) $f_X(x)$ if for every interval $[a, b]$,

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$$

(see Figure 5.4). A random variable which has a probability density function is called a *continuous* random variable.³

The probability density function of a continuous random variable X can be interpreted as the “idealized histogram” of the values of X if we repeat the experiment many many times. More specifically, suppose we repeat the experiment n times, and plot the histogram of the values of X in n repeated experiments with bins of width ε . If ε is chosen small and n is chosen large, then the histogram approximates the pdf $f_X(x)$.

Ⓚ What is the total area under the pdf of a continuous random variable?

Ⓐ The total area is 1. Namely, the total area is the probability that $-\infty < X < \infty$, which is 1.

²The experiments are simulated on a computer using R.

³... or more accurately, an *absolutely continuous* random variable.

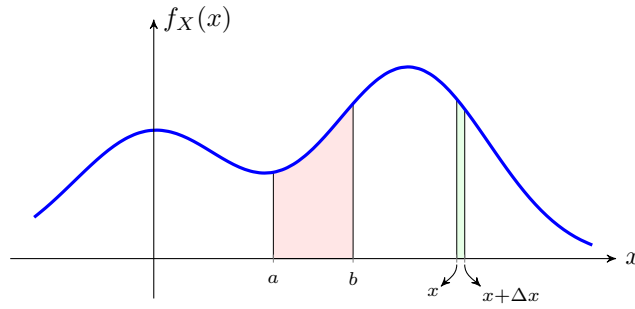


Figure 5.4: The probability density function of a continuous random variable X is an “idealized histogram”. The area under f_X between a and b is the probability that $a \leq X \leq b$.



Relationship between the pdf and the cdf. The two functions $f_X(x)$ and $F_X(x)$ carry the same information about a continuous random variable X : if we know one of them, we can find the other.

Q If we know the pdf of a continuous random variable X , how can we find its cdf?

A

$$F_X(x) = \mathbb{P}(-\infty < X \leq x) = \int_{-\infty}^x f_X(y) dy .$$

Q If we know the cdf of a continuous random variable X , how can we find its pdf?

A

Suppose we know the cdf F_X and we want to find the value of the pdf $f_X(x)$ for some value x . Consider another point $x + \Delta x$, where Δx is small. Then, the area under f_X between x and $x + \Delta x$ can be approximated by the area of a rectangle with width Δx and height $f_X(x)$, hence

$$f_X(x)\Delta x \approx \mathbb{P}(x \leq X \leq x + \Delta x) = F_X(x + \Delta x) - F_X(x) .$$

Therefore,

$$f_X(x) \approx \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} .$$

Sending $\Delta x \rightarrow 0$, the approximation becomes accurate and we obtain

$$f_X(x) = \frac{dF_X}{dx}(x) .$$



Shape of the cdf (continuous RVs). Let $F(x)$ be the cdf of a *continuous* random variable X .⁴ Then,

- (i) $F(x)$ is non-decreasing.
- (ii) $F(x)$ is a continuous function.
- (iii) $F(x) \rightarrow 0$ as $x \rightarrow -\infty$.
- (iv) $F(x) \rightarrow 1$ as $x \rightarrow +\infty$.

Example 5.2.2 (An exponential RV). An *exponential random variable* with parameter $\lambda > 0$ is a continuous random variable T with cdf

$$F_T(t) := \begin{cases} 1 - e^{-\lambda t} & \text{if } t \geq 0, \\ 0 & \text{if } t < 0 \end{cases}$$

(see Figure 5.5a). The distribution of an exponential random variable is called an *exponential distribution*. The parameter λ is called the *rate* of the distribution. The pdf of T is simply the derivative of its cdf, which is

$$f_T(t) := \begin{cases} \lambda e^{-\lambda t} & \text{if } t > 0, \\ 0 & \text{if } t < 0 \end{cases}$$

(see Figure 5.5b). Note that $f_T(t)$ is not defined at $t = 0$, because $F_T(t)$ is not differentiable at that point. This however causes no problem because we only care about the integrals of $f_T(t)$, which are still meaningful if $f_T(t)$ is undefined for a “negligible” set of points.

An exponential random variable is a reasonable model for the arrival time of a “rare” event, for instance:

⁴Compare with the shape of the cdf for discrete random variables.

- time until you receive a spam email in your mailbox,
- time until the next customer arrives at a shop,
- time until the next visitor visits a website,
- time until a block of radioactive material emits a particle,
- ...

In these cases, the parameter λ is interpreted as the average rate of receiving spams, the average rate of the arrival of customers, and so on.

Observe that, if T is an exponential random variable with parameter λ , then

$$\mathbb{P}(T > t) = 1 - F_T(t) = \begin{cases} e^{-\lambda t} & \text{if } t \geq 0, \\ 1 & \text{if } t < 0. \end{cases}$$

This is a somewhat simpler characterization of the distribution of T . ○

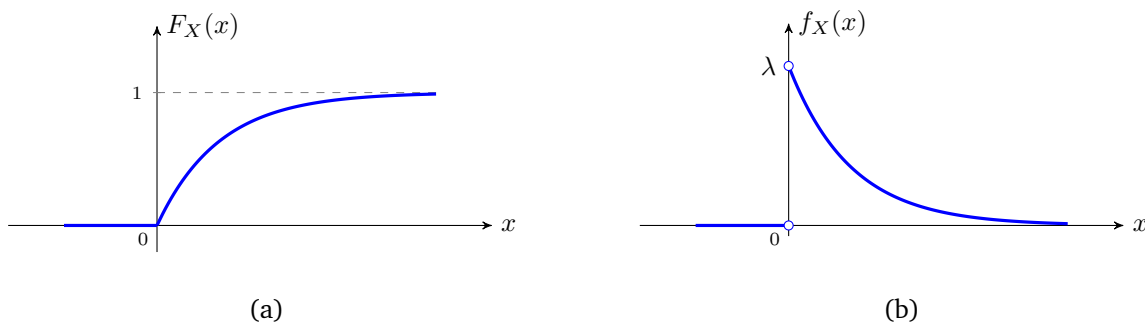


Figure 5.5: The distribution of an exponential random variable with rate λ . (a) cdf (b) pdf

Exponential random variables can be thought of as continuous analogues of geometric random variables. We will later talk about this analogy in more details.



Exponential RVs are memoryless. Exponential random variables are *memoryless* in a sense similar to geometric random variables. Namely, let T be an exponential random variable with rate λ . Then, for every $s, t \geq 0$, we have

$$\begin{aligned} \mathbb{P}(T > s + t \mid T > s) &= \frac{\mathbb{P}(T > s + t \text{ and } T > s)}{\mathbb{P}(T > s)} = \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= \mathbb{P}(T > t). \end{aligned}$$

In other words, the conditional distribution of $T - s$ given $T > s$ is the same as the distribution of T . For instance, if T stands for the duration from time 0 until you receive the first spam in your mail box, then the memorylessness of T simply says the following:

- If we learn that until time s , no spam email has arrived, then the duration from time s until the first spam email arrives will again be exponentially distributed with the same rate.

Example 5.2.3 (A normal RV). Recall the bell-shaped histograms of the weights and heights of 18-year-old individuals in Hong-Kong (Section 2.1.2). The normal distribution is the mathematical idealization of such histograms.

The *standard normal distribution* is the continuous distribution described by the pdf

$$\varphi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

(see Figure 5.6b). Observe that this is a symmetric unimodal distribution. A *standard normal random variable* is a continuous random variable with standard normal distribution.

The cdf of the standard normal distribution is the integral of its pdf, that is,

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}x^2} dx$$

(see Figure 5.6a). Unfortunately, the latter integral does not have a closed form in terms of elementary functions.⁵ In practice, one can use a computer to numerically calculate the value of $\Phi(z)$ for any input, with high accuracy. One can also numerically calculate the inverse of Φ (i.e., given $0 < p < 1$, find value z such that $\Phi(z) = p$).⁶

Q In Figure 5.6a, what is the value of Φ at the point where the graph of Φ crosses the vertical axis?

A $1/2$. The value in question is $\Phi(0) = \mathbb{P}(Z \leq 0)$. This is the area under the pdf to the left of the vertical axis. Since the pdf is symmetric around the vertical axis (it is an even function), the area is $1/2$. $\circ$

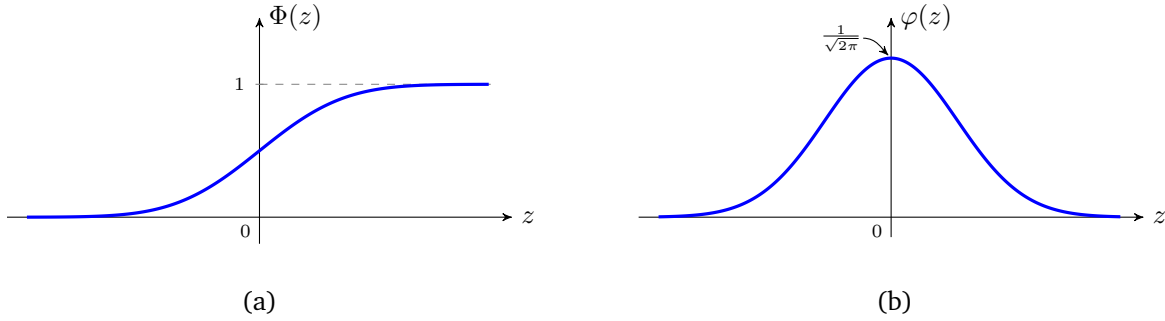


Figure 5.6: The distribution of a standard normal random variable. (a) cdf (b) pdf

 *Exercise.* Verify that $f_Z(z) := \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$ is indeed a probability density function, in particular,

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}.$$

Hint: Let $S := \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx$, and note that $S^2 = \iint_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2+y^2)} dx dy$. Use polar coordinates.

5.3 Expected value of a random variable: part II

Recall that the expected value of a random variable X is the “idealized mean” of the measurements of X in many repeated experiments. Namely, if we repeat the experiment n times (n large), then

$$\mathbb{E}[X] \approx \frac{x_1 + x_2 + \cdots + x_n}{n},$$

where x_k stands for the value of X in the k 'th experiment. For discrete random variables, this lead to the expression

$$\mathbb{E}[X] = \sum_a \mathbb{P}(X = a)a,$$

in terms of the probability mass function.

For continuous random variables, we can derive a similar expression for the expected value as follows. Divide the real axis into small intervals $[a_i, a_{i+1})$ of common length Δa (see Figure 5.7). As usual, let $N_n(a_i \leq X < a_{i+1})$ denote the number of experiments in which the value of X falls within the interval $[a_i, a_{i+1})$. Then,

$$\begin{aligned} \mathbb{E}[X] &\approx \frac{x_1 + x_2 + \cdots + x_n}{n} \\ &\approx \frac{1}{n} \sum_i N_n(a_i \leq X < a_{i+1}) \cdot a_i \\ &\approx \sum_i a_i \cdot \mathbb{P}(a_i \leq X < a_{i+1}) \\ &= \sum_i a_i \int_{a_i}^{a_{i+1}} f(x) dx \\ &\approx \int_{-\infty}^{\infty} x f(x) dx. \end{aligned}$$

⁵For more information, see [Liouville's theorem \(Wikipedia\)](#) and the references therein.

⁶Some programming languages have standard routines or libraries for calculating the normal cdf and its inverse. For instance, in R, the normal cdf can be calculated using the function `pnorm` and its inverse by `qnorm`. In Python, the module `scipy.stats` has functions `norm.cdf` and `norm.ppf` for computing the normal cdf and its inverse.

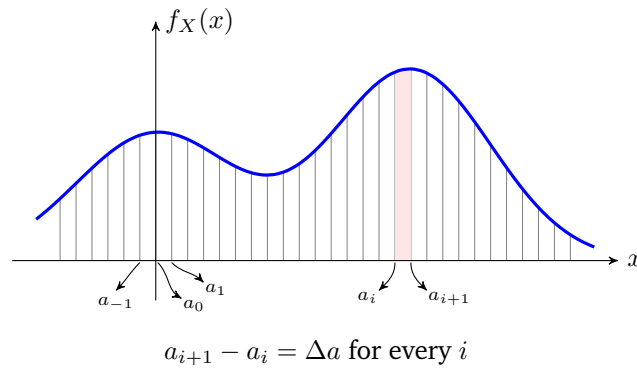


Figure 5.7: The probability density function of a continuous random variable X . The values of x are partitioned into intervals of length Δa .

For the second identity, we have approximated all the values within interval $[a_i, a_{i+1})$ with a_i , for the third identity, we have used the fact that the ratio $N_n(a_i \leq X < a_{i+1})/n$ approximates $\mathbb{P}(a_i \leq X < a_{i+1})$, and for the last identity, we have again used the fact that $a_i \approx x$ for $x \in [a_i, a_{i+1})$. Choosing Δa smaller and smaller, and n larger and larger, we obtain more and more accurate approximations, and in the limit, we get an exact equality.



Terminology. The *expected value* of a continuous random variable X is the quantity

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

The variance and standard deviation of X are defined via expectation as in the case of discrete random variables.

As in the discrete case, the expected value of a function $h(X)$ of a continuous random variable X can be calculated as

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

In other words, in order to find the expected value of $h(X)$, we do not need to explicitly know the distribution of $h(X)$; we can instead use the distribution of X .

Ⓚ Why does the latter expression for $\mathbb{E}[h(X)]$ hold?

Ⓐ The justification is similar to the derivation of the expression for $\mathbb{E}[X]$, except that we group the terms based on the values of X rather than the values of $h(X)$.

Example 5.3.1 (A uniform RV). Let X be a continuous random variable which is uniformly distributed over an interval $[a, b]$ (compare with Examples 5.1.1 and 5.2.1). The pdf and the cdf of X are depicted in Figure 5.8.

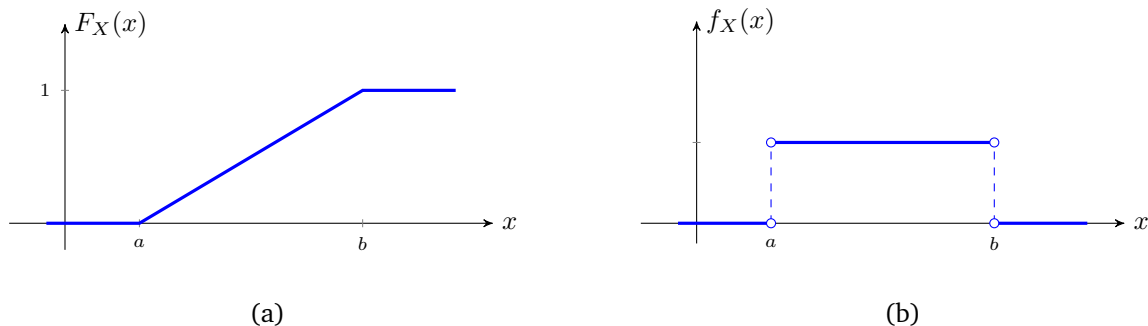


Figure 5.8: The distribution of a uniform RV over an interval $[a, b]$. (a) cdf (b) pdf

Ⓚ What is value of $f_X(x)$ for $x \in (a, b)$?

Ⓐ $1/(b - a)$. Total area under the pdf must be 1.

Ⓚ What is the expected value of X ?

A1

$$\begin{aligned}\mu_X &:= \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx \\ &= \int_a^b x \cdot \frac{1}{b-a} \, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} .\end{aligned}$$

A2 $\frac{a+b}{2}$. The distribution is symmetric about the mid-point between a and b , which is $x = (a+b)/2$.

Q What is the variance of X ?

A

$$\begin{aligned}\mathbb{V}\text{ar}[X] &:= \mathbb{E}[(X - \mu_X)^2] \\ &= \int_{-\infty}^{\infty} \left(x - \frac{a+b}{2}\right)^2 f_X(x) \, dx \\ &= \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} \, dx \\ &= \frac{1}{3(b-a)} \left[\left(\frac{b-a}{2}\right)^3 - \left(\frac{b-a}{2}\right)^3 \right] = \frac{(b-a)^2}{12} .\end{aligned}$$

It follows that

$$\mathbb{S}\mathbb{D}[X] = \frac{1}{2\sqrt{3}}(b-a) .$$

In particular, the standard deviation of X is proportional to the length of the interval $[a, b]$. This is consistent with the interpretation of the standard deviation as a measure of the spread of the distribution. $\circ$



Some basic facts about the expectation and variance. The following properties of expectation and variance of discrete random variables extend to continuous random variables (and indeed any type of random variables).

- (expectation: shifting and scaling) If X is a random variable and $a, b \in \mathbb{R}$ arbitrary numbers, then

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b .$$

- (expectation: linearity) If X and Y are two random variable, then

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] .$$

- (expectation: monotonicity) If X and Y are random variables where $X \leq Y$, then

$$\mathbb{E}[X] \leq \mathbb{E}[Y] .$$

- (expectation: product of independent RVs) If X and Y are two independent random variables, then

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] .$$

- (variance: shifting and scaling) If X is a random variable and $a, b \in \mathbb{R}$ arbitrary numbers, then

$$\begin{aligned}\mathbb{V}\text{ar}[aX + b] &= a^2 \mathbb{V}\text{ar}[X] , \\ \mathbb{S}\mathbb{D}[aX + b] &= |a| \mathbb{S}\mathbb{D}[X] .\end{aligned}$$

- (variance: sum of independent RVs) If X and Y are two independent random variables, then

$$\mathbb{V}\text{ar}[X + Y] = \mathbb{V}\text{ar}[X] + \mathbb{V}\text{ar}[Y] .$$

Example 5.3.2 (An exponential RV). Let T be an exponential random variable with rate λ , as in Example 5.2.2.

Q What is the expected value of T ?

A

$$\begin{aligned}
 \mathbb{E}[T] &= \int_{-\infty}^{\infty} t f_T(t) dt \\
 &= \int_0^{\infty} t \cdot \lambda e^{-\lambda t} dt \\
 &= \overset{0}{-te^{-\lambda t}} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt \quad (\text{integration by parts}) \\
 &= \frac{e^{-\lambda t}}{-\lambda} \Big|_0^{\infty} = \frac{0 - 1}{-\lambda} = \frac{1}{\lambda}.
 \end{aligned}$$

In words, the mean of an exponential random variable is the inverse of its rate.

Q What is the variance of T ?

A We use the identity $\text{Var}[T] = \mathbb{E}[T^2] - (\mathbb{E}[T])^2$. We have,

$$\begin{aligned}
 \mathbb{E}[T^2] &= \int_{-\infty}^{\infty} t^2 f_T(t) dt \\
 &= \int_0^{\infty} t^2 \cdot \lambda e^{-\lambda t} dt \\
 &= \overset{0}{-t^2 e^{-\lambda t}} \Big|_0^{\infty} + \int_0^{\infty} 2te^{-\lambda t} dt \quad (\text{integration by parts}) \\
 &= \frac{2}{\lambda} \underbrace{\int_0^{\infty} t \cdot \lambda e^{-\lambda t} dt}_{\mathbb{E}[T]} = \frac{2}{\lambda^2}.
 \end{aligned}$$

Therefore,

$$\text{Var}[T] = \mathbb{E}[T^2] - (\mathbb{E}[T])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

It follows that

$$\text{SD}[T] = \frac{1}{\lambda}.$$

In particular, the standard deviation of an exponential random variable is inversely proportional to its rate: the higher the rate, the lower the spread. ○

Example 5.3.3 (A standard normal RV). Let Z be a standard normal random variable, hence Z have pdf

$$\varphi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

(see Example 5.2.3 and Figure 5.6b).

Q What is the expected value of Z ?

A1 0. Note that the distribution of Z is symmetric about $z = 0$.

A2

$$\mathbb{E}[Z] = \int_{-\infty}^{\infty} z \varphi(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{ze^{-\frac{1}{2}z^2}}_{\text{an odd function}} dz = 0$$

Q What is the variance of Z ?

[A] Since $\mathbb{E}[Z] = 0$, we have

$$\begin{aligned}
 \text{Var}[Z] &= \mathbb{E}[Z^2] = \int_{-\infty}^{\infty} z^2 \varphi(z) \, dz \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} \, dz \\
 &= \frac{1}{\sqrt{2\pi}} \left[-ze^{-\frac{1}{2}z^2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \, dz \right] \quad \left(\text{integration by parts with } \begin{cases} u(z) := z \\ v(z) := -e^{-\frac{1}{2}z^2} \end{cases} \right) \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \, dz \\
 &= \int_{-\infty}^{\infty} \varphi(z) \, dz = 1 .
 \end{aligned}$$

○



Scaling and shifting standard normal RVs. A scaled and shifted version of a *standard* normal random variable is called a *normal random variable*. More specifically, a random variable X is a normal random variable if it can be represented as $X = \sigma Z + \mu$ where Z is a standard normal random variable and σ and μ are two real numbers. The distribution of a normal random variable is called a *normal distribution*.

Observe that in this case,

$$\begin{aligned}
 \mathbb{E}[X] &= \mathbb{E}[\sigma Z + \mu] = \sigma \mathbb{E}[Z] + \mu = \sigma \cdot 0 + \mu = \mu , \\
 \text{Var}[X] &= \text{Var}[\sigma Z + \mu] = \sigma^2 \text{Var}[Z] = \sigma^2 \cdot 1 = \sigma^2 .
 \end{aligned}$$

Therefore, the two parameters μ and σ are simply the expectation and the standard deviation of X . Sometimes, μ is referred to as the *location* parameter, and σ is referred to as the *scale* parameter of the normal distribution.



Notation. The normal distribution with mean μ and variance σ^2 is denoted by $N(\mu, \sigma^2)$. With this notation, the standard normal distribution is represented by $N(0, 1)$. We write $X \sim N(\mu, \sigma^2)$ to indicate that X is a normal random variable with mean μ and variance σ^2 .



Independence of RVs. Recall that the independence of two discrete random variables X and Y could be expressed with the following condition:

- For every two values x, y , the events $\{X = x\}$ and $\{Y = y\}$ are independent.

This formulation is not adequate for non-discrete random variables, so we need a more general formulation.

In general, we can say that two random variables X and Y (of any type) are (statistically) *independent* if

- Every two events of the form $\{X \in A\}$ and $\{Y \in B\}$ are independent.

The independence of more than two random variables can be formulated analogously.



Exercise. Argue that, for discrete random variables, the above two formulations of independence are equivalent.