Probability and Statistics (IT302)

19th August 2020 (Wednesday) 11:15 AM-11:45 AM

Class 7

Introduction to Probability Density Function/Density Function

- Let us discuss a Random Variable whose values are the heights of all people over 21 years of age. Between any two values, say 163.5 and 164.5 centimeters, or even 163.99 and 164.01 centimeters, there are an infinite number of heights, one of which is 164 centimeters. The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters that you cannot humanly measure the difference is remote, and thus assign a probability of 0 to the event.

- This is not the case, however, if we talk about the probability of selecting a person who is at least 163 centimeters but not more than 165 centimeters tall. Now we are dealing with an interval rather than a point value of our Random Variable.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall.

Introduction to Probability Density Function/Density Function Contd.

We shall concern ourselves with computing probabilities for various intervals of Continuous Random Variables such as $P(a < X < b)$, $P(W \geq c)$, and so forth.

Note that when X is continuous,

$$P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b).$$

That is, it does not matter whether we include an endpoint of the interval or not. This is not true, though, when X is discrete.

Although the Probability Distribution of a Continuous Random Variable cannot be presented in tabular form, it can be stated as a formula. Such a formula would necessarily be a function of the numerical values of the Continuous Random Variable X and as such will be represented by the functional notation $f(x)$.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Introduction to Probability Density Function/Density Function Contd.

In dealing with Continuous Variables, $f(x)$ is usually called the **Probability Density Function**, or simply the **Density Function**, of X. Since X is defined over a continuous sample space, it is possible for $f(x)$ to have a finite number of discontinuities. However, most Density Functions that have practical applications in the analysis of statistical data are continuous and their graphs may take any of several forms, some of which are shown in Figure 3.4.

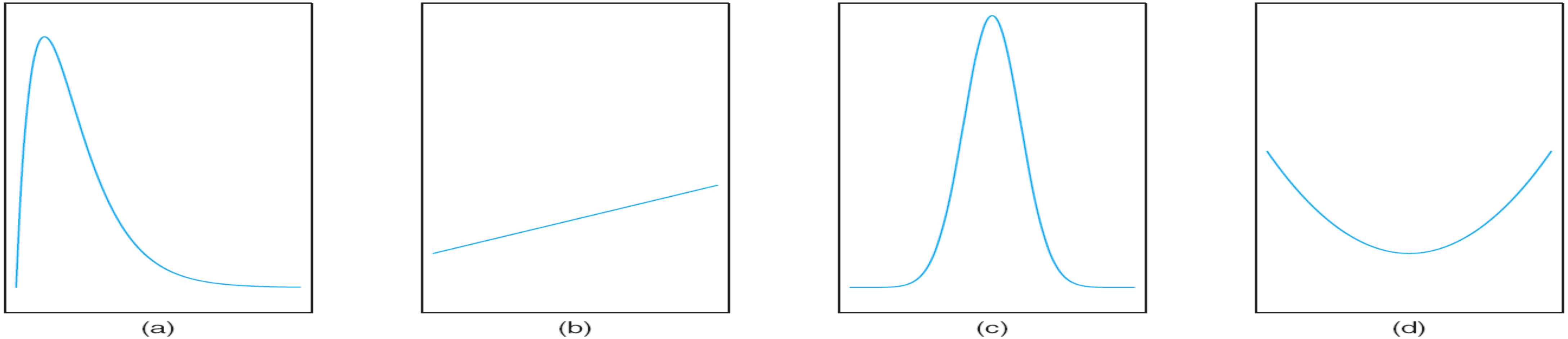


Figure 3.4: Typical density functions.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Introduction to Probability Density Function/Density Function Contd.

bounded by the x axis is equal to 1 when computed over the range of X for which $f(x)$ is defined. Should this range of X be a finite interval, it is always possible to extend the interval to include the entire set of real numbers by defining $f(x)$ to be zero at all points in the extended portions of the interval. In Figure 3.5, the probability that X assumes a value between a and b is equal to the shaded area under the density function between the ordinates at $x = a$ and $x = b$, and from integral calculus is given by

$$P(a < X < b) = \int_a^b f(x)\, dx.$$

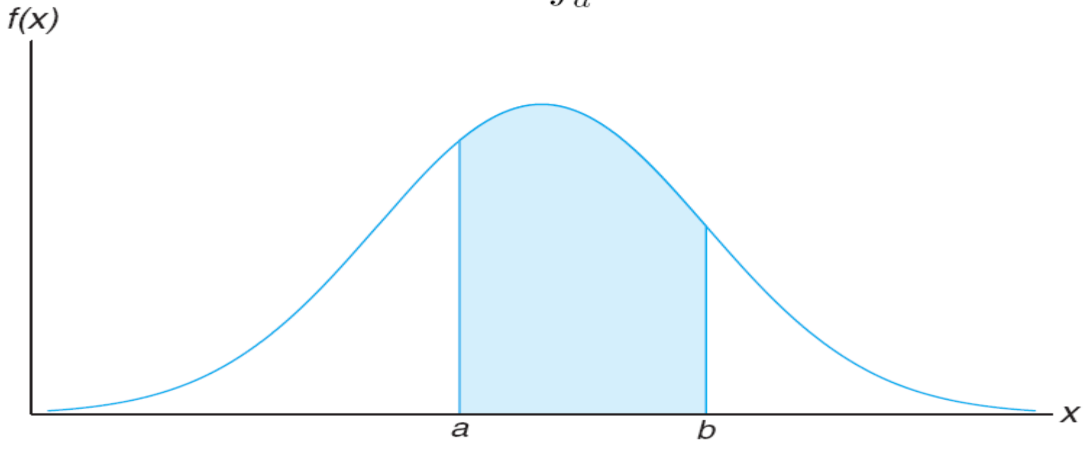


Figure 3.5: $P(a < X < b)$.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Probability Density Function

Definition 3.6

The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall, PP. 84

Probability Density Function Example

Example 3.11: Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

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

.

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

***Solution*:** We use Definition 3.6.

(a) Obviously, $f(x) \geq 0$. To verify condition 2 in Definition 3.6, we have

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1.$$

(b) Using formula 3 in Definition 3.6, we obtain

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_0^1 = \frac{1}{9}.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Cumulative Distribution Function

Definition 3.7: The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt, \quad \text{for } -\infty < x < \infty.$$

As an immediate consequence of Definition 3.7, one can write the two results

$$P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx},$$

if the derivative exists.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Cumulative Distribution Function Example

Example 3.12: For the density function of Example 3.11, find $F(x)$, and use it to evaluate $P(0 < X \leq 1)$.

***Solution*:** For $-1 < x < 2$,

$$F(x) = \int_{-\infty}^{x} f(t)\ dt = \int_{-1}^{x} \frac{t^2}{3} dt = \left. \frac{t^3}{9} \right|_{-1}^{x} = \frac{x^3 + 1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

The cumulative distribution function $F(x)$ is expressed in Figure 3.6. Now

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9},$$

which agrees with the result obtained by using the density function in Example 3.11.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Cumulative Distribution Function Example

Example 3.13: The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \le y \le 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $F(y)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

Solution: For $2b/5 \le y \le 2b$,

$$F(y) = \int_{2b/5}^{y} \frac{5}{8b}\,dy = \left.\frac{5t}{8b}\right|_{2b/5}^{y} = \frac{5y}{8b} - \frac{1}{4}.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Cumulative Distribution Function Example Contd.

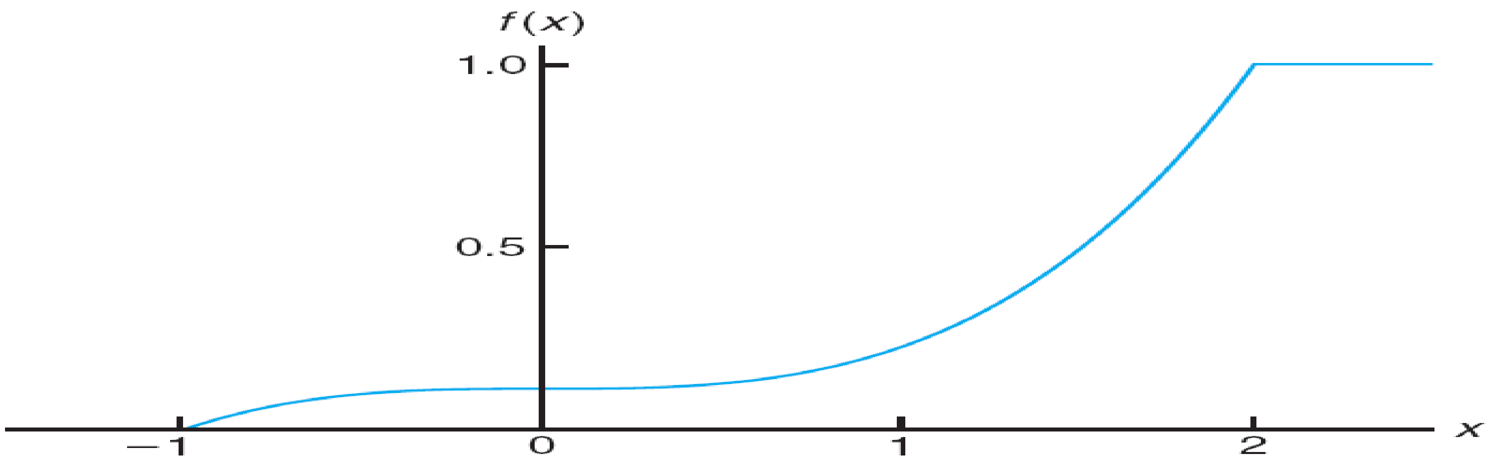


Figure 3.6: Continuous cumulative distribution function.

Thus,

$$F(y) = \begin{cases} 0, & y < \frac{2}{5}b, \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5}b \le y < 2b, \\ 1, & y \ge 2b. \end{cases}$$

To determine the probability that the winning bid is less than the preliminary bid estimate b, we have

$$P(Y \le b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Exercise 3.7

The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a Continuous Random Variable *X* that has the density function

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

Find the probability that over a period of one year, a family runs their vacuum cleaner

(a) less than 120 hours; (b) between 50 and 100 hours.

Solution

$$\text{(a)}\ \ P(X < 1.2) = \int_0^1 x\, dx + \int_1^{1.2} (2 - x)\, dx = \left.\frac{x^2}{2}\right|_0^1 + \left.\left(2x - \frac{x^2}{2}\right)\right|_1^{1.2} = 0.68.$$

$$\text{(b)}\ \ P(0.5 < X < 1) = \int_{0.5}^1 x\, dx = \left.\frac{x^2}{2}\right|_{0.5}^1 = 0.375.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Exercise 3.9

The proportion of people who respond to a certain mail-order solicitation is a continuous random variable X that has the density function

$$f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Show that $P(0 < X < 1) = 1$.

(b) Find the probability that more than 1/4 but fewer than 1/2 of the people contacted will respond to this type of solicitation.

Solution

(a) $P(0 < X < 1) = \int_0^1 \frac{2(x+2)}{5}\, dx = \left.\frac{(x+2)^2}{5}\right|_0^1 = 1.$

(b) $P(1/4 < X < 1/2) = \int_{1/4}^{1/2} \frac{2(x+2)}{5}\, dx = \left.\frac{(x+2)^2}{5}\right|_{1/4}^{1/2} = 19/80.$

Exercise 3.13

The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Construct the cumulative distribution function of X.

Solution

The c.d.f. of X is

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ 0.41, & \text{for } 0 \le x < 1, \\ 0.78, & \text{for } 1 \le x < 2, \\ 0.94, & \text{for } 2 \le x < 3, \\ 0.99, & \text{for } 3 \le x < 4, \\ 1, & \text{for } x \ge 4. \end{cases}$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Exercise 3.14

The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \geq 0. \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders
(a) using the cumulative distribution function of *X*;
(b) using the probability density function of *X*.

Solution

(a) $P(X < 0.2) = F(0.2) = 1 - e^{-1.6} = 0.7981;$

(b) $f(x) = F'(x) = 8e^{-8x}$. Therefore, $P(X < 0.2) = 8\int_0^{0.2} e^{-8x}\, dx = -e^{-8x}\Big|_0^{0.2} = 0.7981.$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Exercise 3.18

A Continuous Random Variable X that can assume values between $x = 2$ and $x = 5$ has a density function given by $f(x) = 2(1 + x)/27$. Find

(a) $P(X < 4)$;

(b) $P(3 \leq X < 4)$.

Solution

$$\text{(a)}\quad P(X < 4) = \int_2^4 \frac{2(1+x)}{27}\, dx = \left.\frac{(1+x)^2}{27}\right|_2^4 = 16/27.$$

$$\text{(b)}\quad P(3 \leq X < 4) = \int_3^4 \frac{2(1+x)}{27}\, dx = \left.\frac{(1+x)^2}{27}\right|_3^4 = 1/3.$$

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Exercise 3.20

A Continuous Random Variable X that can assume values between $x = 2$ and $x = 5$ has a density function given by $f(x) = 2(1 + x)/27$. Find $F(x)$. Use it to evaluate $P(3 \leq X < 4)$.

Solution

$$F(x) = \frac{2}{27}\int_2^x (1+t)\,dt = \frac{2}{27}\left(t + \frac{t^2}{2}\right)\Bigg|_2^x = \frac{(x+4)(x-2)}{27},$$

$$P(3 \leq X < 4) = F(4) - F(3) = \frac{(8)(2)}{27} - \frac{(7)(1)}{27} = \frac{1}{3}.$$

Exercise 3.21

The time to failure in hours of an important piece of electronic equipment used in a manufactured DVD player has the density function

$$f(x) = \begin{cases} \frac{1}{2000}\exp(-x/2000), & x \geq 0, \\ 0, & x < 0. \end{cases}$$

a) Find $F(x)$.
b) Determine the probability that the component (and thus the DVD player) lasts more than 1000 hours before the component needs to be replaced.
c) Determine the probability that the component fails before 2000 hours.

Solution

(a) For $x \geq 0$, $F(x) = \int_0^x \frac{1}{2000}\exp(-t/2000)\,dt = -\exp(-t/2000)|_0^x$ $= 1 - \exp(-x/2000)$. So

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-x/2000), & x \geq 0. \end{cases}$$

(b) $P(X > 1000) = 1 - F(1000) = 1 - [1 - \exp(-1000/2000)] = 0.6065$.
(c) $P(X < 2000) = F(2000) = 1 - \exp(-2000/2000) = 0.6321$.

Source : Probability & Statistics for Engineers & Scientists, by Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, 9th Edition, Prentice Hall

Additional Material Continuous Random Variable

- A Random Variable X is **continuous** if possible values comprise either a single interval on the number line or a union of disjoint intervals.

- Example: If in the study of the ecology of a lake, X, the Random Variable may be depth measurements at randomly chosen locations.

- Then X is a continuous Random Variable. The range for X is the minimum depth possible to the maximum depth possible.
- In principle variables such as height, weight, and temperature are continuous, in practice the limitations of our measuring instruments restrict us to a discrete (though sometimes very finely subdivided) world.

- However, continuous models often approximate real-world situations very well, and continuous mathematics (calculus) is frequently easier to work with than mathematics of discrete variables and distributions.

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material Probability Distributions for Continuous Variables

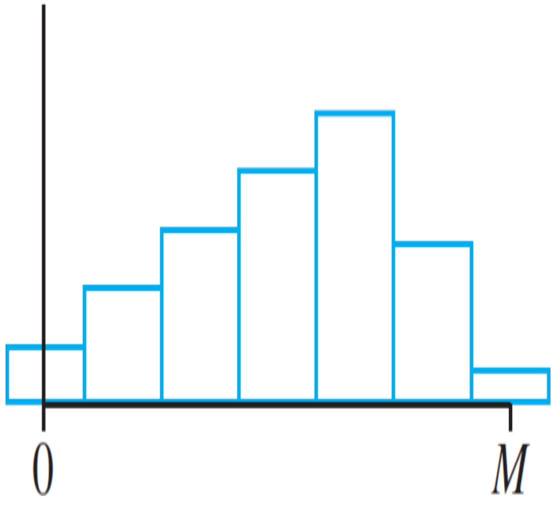
- Suppose the variable X of interest is the depth of a lake at a randomly chosen point on the surface.

- Let M = the maximum depth (in meters), so that any number in the interval $[0, M]$ is a possible value of X.

- If we "discretize" X by measuring depth to the nearest meter, then possible values are nonnegative integers less than or equal to M.

- The resulting discrete distribution of depth can be pictured using a probability histogram.

- If we draw the histogram so that the area of the rectangle above any possible integer k is the proportion of the lake whose depth is (to the nearest meter) k, then the total area of all rectangles is 1:

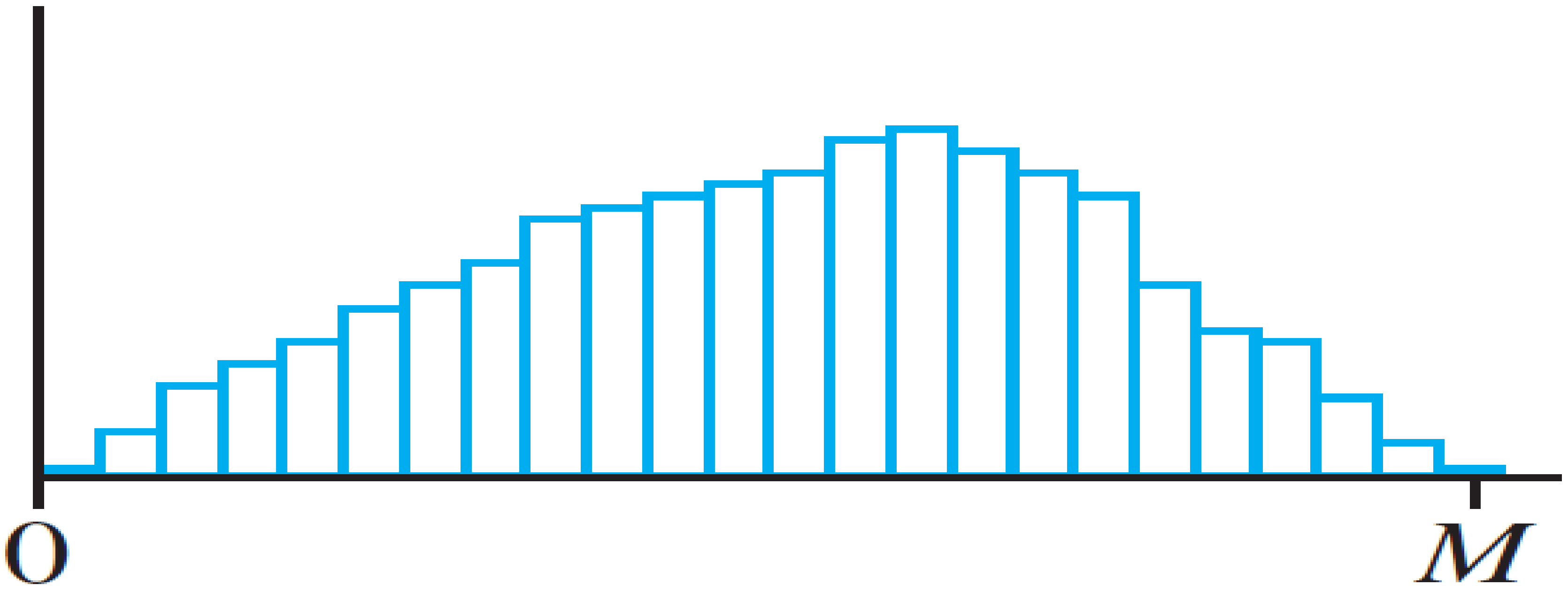


Probability histogram of depth measured to the nearest meter

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material Probability Distributions for Continuous Variables

If depth is measured much more accurately, each rectangle in the resulting probability histogram is much narrower, though the total area of all rectangles is still 1.

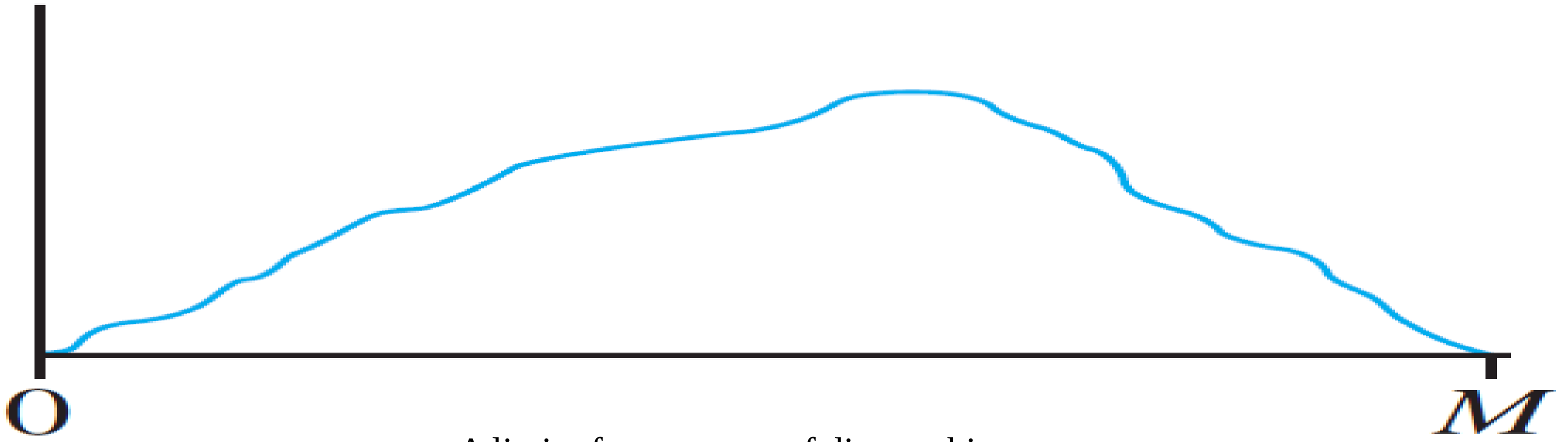


Probability histogram of depth measured to the nearest centimeter

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material Probability Distributions for Continuous Variables

- If we continue in this way to measure depth more and more finely, the resulting sequence of histograms approaches a smooth curve.
- Because for each histogram the total area of all rectangles equals 1, the total area under the smooth curve is also 1.



A limit of a sequence of discrete histograms

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material Probability Distributions for Continuous Variables

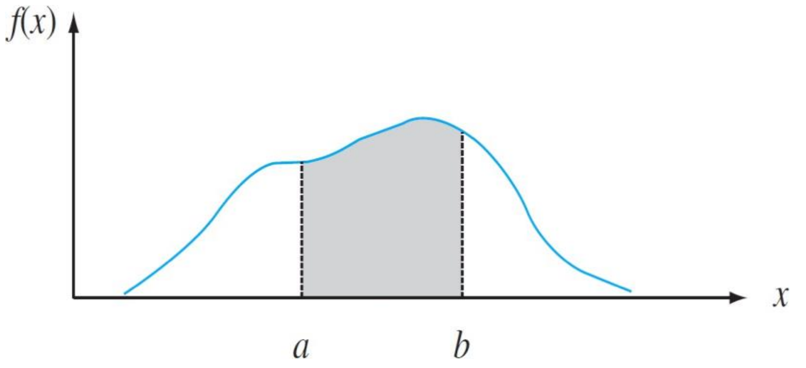
<u>Definition</u>

Let X be a Continuous Random Variable. Then a **Probability Distribution** or **<u>Probability Density Function</u>** (PDF) of X is a function $f(x)$ such that for any two numbers a and b with $a \leq b$, we have

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

The probability that X is in the interval $[a, b]$ can be calculated by integrating the PDF of the Random Variable X.

The probability that X takes on a value in the interval $[a, b]$ is the area above this interval and under the graph of the density function:



$P(a \leq X \leq b)$ = the area under the density curve between a and b

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material Probability Distributions for Continuous Variables

For $f(x)$ to be a legitimate PDF, it must satisfy the following two conditions:

1. $f(x) \geq 0$ for all x

2. $\int_{-\infty}^{\infty} f(x)\,dx$ = area under the entire graph of $f(x) = 1$

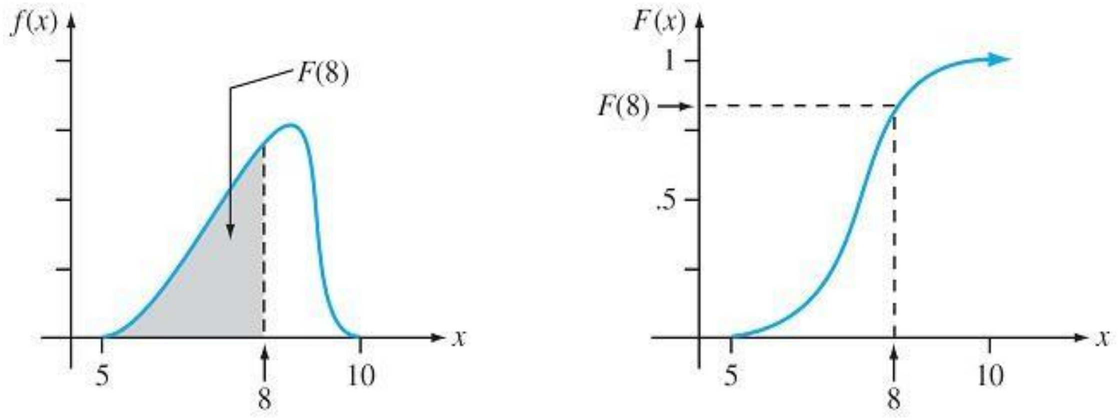
Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material for Cumulative Distribution Function

The **Cumulative Distribution Function** $F(x)$ for a Continuous Random Variable X is defined for every number x by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy$$

For each x, $F(x)$ is the area under the density curve to the left of x. This is illustrated in Figure 4.5, where $F(x)$ increases smoothly as x increases.



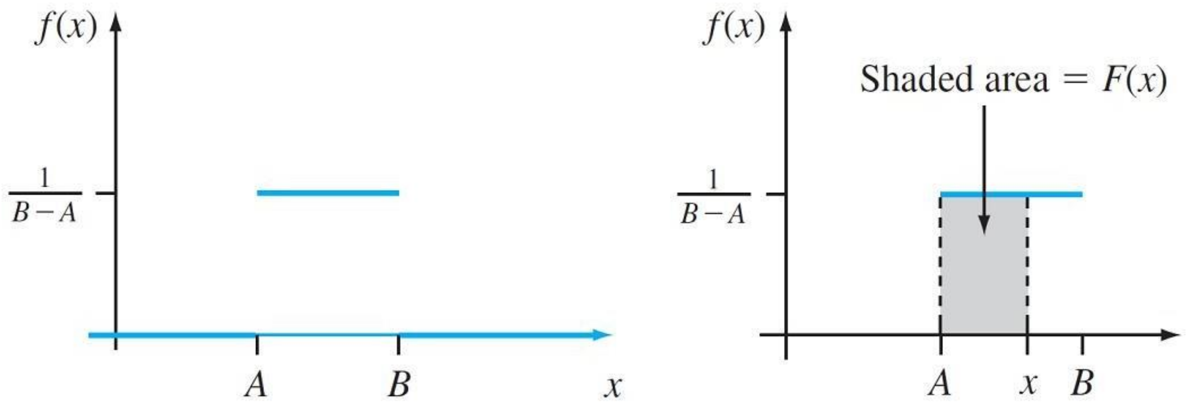
A PDF and associated CDF **Figure 4.5**

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material for Cumulative Distribution Function

Example

Let X, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The density function is shown in Figure 4.6.



The PDF for a uniform distribution **Figure 4.6**

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material for Cumulative Distribution Function

Example Contd.

For $x < A$, $F(x) = 0$, since there is no area under the graph of the density function to the left of such an x.

For $x \geq B$, $F(x) = 1$, since all the area is accumulated to the left of such an x. Finally for $A \leq x \leq B$,

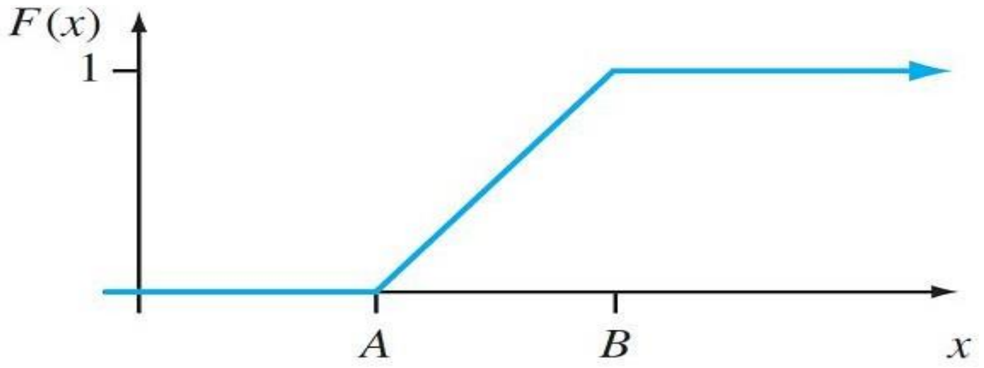
$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{A}^{x} \frac{1}{B - A} dy = \frac{1}{B - A} \cdot y \Big|_{y=A}^{y=x} = \frac{x - A}{B - A}$$

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf

Additional Material for Cumulative Distribution Function

The entire CDF is $$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x - A}{B - A} & A \le x < B \\ 1 & x \ge B \end{cases}$$

The graph of this CDF appears in Figure 4.7.



The CDF for a uniform distribution **Figure 4.7**

Source : https://www.colorado.edu/amath/sites/default/files/attached-files/ch4.pdf