Lesson 3.1

Probability Density Function and Cumulative Distribution Function of Continuous Random Variables

Learning Outcomes

At the end of the lesson, students must be able to

1. Explain the definition of probability density function (PDF) and cumulative distribution function (CDF) of continuous random variables,
2. Derive the PDF from the CDF and vice versa, and
3. Compute probabilities associated with a continuous random variable using either its PDF or CDF.

Introduction

The last module dealt with discrete random variables. A discrete random variable Y can assume a finite or (at most) a countable number of values. The probability mass function (pmf) of a discrete random variable

$$P(Y = y) = p_Y(y)$$

specifies how to assign probability to each support point $y \in D$, a countable set.

A continuous random variable takes on an (uncountable) infinite number of possible values. Some examples of continuous random variables are: the amount of rain (mm) that falls in a randomly selected typhoon, the weight (kg) of a randomly selected employee, and the length of time (min) to finish an exam of a randomly selected student.

For continuous random variables, the probability that Y takes on any particular value y is 0. That is, finding $P(Y = y)$ for a continuous random variable Y is not going to work. Instead, we'll need to find the probability that Y falls in some interval (a, b), that is, we'll need to find $P(a < Y < b)$. We'll do that using a probability density function (PDF).

Definitions of the PDF and CDF

Definition:

The probability density function of a continuous random variable Y with domain D is an integrable function $f_Y(y)$ satisfyng the following:

1. $f_Y(y)$ is non-negative everywhere in the domain, that is $f_Y(y) \geq 0,\ \forall y \in D$.
2. The area under the curve $f_Y(y)$ in the domain is 1, that is,

$$\int_{\forall y \in D} f_Y(y)\, dy = 1$$

Remarks:

1. Notice that the definition for the PDF of a continuous random variable differs from the definition for the pmf of a discrete random variable by simply changing the summations that appeared in the discrete case to integrals in the continuous case.
2. For a continuous random variable, the probability that y belongs to an interval I is given by

$$P(y \in I) = \int_{\forall y \in I} f_Y(y)\, dy$$

Example 1:

Let Y be a continuous random variable whose probability density function is

$$f_Y(y) = \begin{cases} 3y^2,\ 0 < y < 1 \\ 0,\ \text{elsewhere} \end{cases}$$

a. Verify that indeed $f_Y(y)$ is a valid PDF.
b. What is the probability that Y falls in the interval (0.5, 1)?
c. What is the probability that $Y = \frac{1}{2}$?

SOLUTION:

a. First we check if $f_Y(y)$ is non-negative over the domain or support (0, 1). Obviously, as y assumes values in the interval (0, 1) $f_Y(y)$ assumes values in the interval (0,3). So the first condition in the definition is meet. Now, it can be shown that

$$\int_0^1 3y^2 \, dy = 1$$

which means that the second condition in the definition is also met. Therefore, $f_Y(y)$ is a valid PDF.

b. The probability that Y falls in the interval (0.5,1) is

$$\begin{aligned} P(0.5 < Y < 1) &= \int_{0.5}^1 3y^2 \, dy \\ &= y^3\Big|_{0.5}^1 \\ &= 0.875 \end{aligned}$$

c. The probability that $Y = \frac{1}{2}$ is

$$\begin{aligned} P(Y = \frac{1}{2}) &= P(\frac{1}{2} < Y < \frac{1}{2}) \\ &= \int_{0.5}^{0.5} 3y^2 \, dy \\ &= y^3\Big|_{0.5}^{0.5} \\ &= 0 \end{aligned}$$

You might recall that the cumulative distribution function is defined for discrete random variables as

$$F_Y(y) = P(Y \leq y) = \sum_{\forall t \leq y} P(Y = t)$$

Again, $F_y(y)$ accumulates all of the probability that Y is less than or equal to y. The cumulative distribution function for continuous random variables is just a straightforward extension of that of the discrete case. All we need to do is replace the summation with an integral.

Definition:

The cumulative distribution function (CDF) of a continuous random variable Y is defined as

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} f_y(t) \, dt$$

Recall that for discrete random variables $F_Y(y)$ is, in general, a non-decreasing step function. For continuous random variables, $F_Y(y)$ is a non-decreasing continuous function.

Relationship between PDF and CDF

Based on the above definition, the CDF is obtained by integrating the PDF. Can we derive the PDF from the CDF?

Suppose Y is a continuous random variable with CDF $F_Y(y)$ and PDF $f_Y(y)$. Then

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\, dt$$

and

$$f_y(y) = \frac{d}{dy} F_Y(y), \text{ provided the derivative exists}$$

Example 2:

Consider the function

$$f_Y(y) = \begin{cases} y^2 + ky, \ 0 \leq y \leq 1 \\ 0, \text{ elsewhere} \end{cases}$$

a. Find the constant k such that $f_Y(y)$ becomes a valid PDF.
b. Derive the CDF of Y.

SOLUTION:

a. We will use the second condition in the definition of the PDF to calculate the value of k.

$$\begin{aligned} \int_{\forall y \in D} f_Y(y)\, dy = 1 &\implies \int_0^1 (y^2 + ky)\, dy = 1 \\ &\implies \frac{y^3}{3} + k\frac{y^2}{2}\Big|_0^1 = 1 \\ &\implies k = \frac{4}{3} \end{aligned}$$

Thus,

$$f_Y(y) = \begin{cases} y^2 + \frac{4}{3}y, \ 0 \leq y \leq 1 \\ 0, \text{ elsewhere} \end{cases}$$

Notice that as Y assumes values in between 0 and 1, the function $f_Y(y) \geq 0$ meaning it satisfies condition 1 of the definition of the PDF.

b. The CDF of Y is determined as follows:

If $y < 0$, then $F_Y(y) = \int_{-\infty}^{y} 0\,dt = 0$.

If $0 \leq y < 1$, then $F_Y(y) = \int_{-\infty}^{0} 0\,dt + \int_{0}^{y} (t^2 + \frac{4}{3}t)\,dt = \frac{y^3}{3} + \frac{2}{3}y^2$

If $y \geq 1$, then $F_Y(y) = \int_{-\infty}^{0} 0\,dt + \int_{0}^{1} (t^2 + \frac{4}{3}t)\,dt + \int_{1}^{\infty} 0\,dt = 1$

Therefore,

$$F_Y(y) = \begin{cases} 0, \ y < 0 \\ \frac{y^3}{3} + \frac{2}{3}y^2, \ 0 \leq y < 1 \\ 1, \ y \geq 1 \end{cases}$$

Remark:

If Y is a continuous random variable with CDF $F_Y(y)$ and PDF $f_Y(y)$, then for any real numbers $a < b$,

$$P(a < Y < b) = P(a \leq Y < b) = P(a < Y \leq b) = P(a \leq Y \leq b)$$

and each equals

$$\int_a^b f_Y(y)\,dy = F_Y(b) - F_Y(a)$$

Example 3:

Suppose Y is a continuous random variable with CDF

$$F_Y(y) = \frac{1}{1 + e^{-y}}, \ -\infty < y < \infty$$

a. Find the PDF of Y.

b. Calculate P(-2 < Y < 2) using the CDF and the PDF.

SOLUTION:

a. The PDF of Y is

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy}\left[\frac{1}{1+e^{-y}}\right] \\ &= \frac{e^{-y}}{(1+e^{-y})^2} \end{aligned}$$

Hence,

$$f_Y(y) = \frac{e^{-y}}{(1+e^{-y})^2},\ -\infty < y < \infty$$

b. Using the CDF

$$P(-2 < Y < 2) = F_Y(2) - F_Y(-2) = \frac{1}{1+e^{-2}} - \frac{1}{1+e^{-(-2)}} \approx 0.762$$

Using the PDF:

$$P(-2 < Y < 2) = \int_{-2}^{2} \frac{e^{-y}}{(1+e^{-y})^2}\, dy \approx 0.762$$

HINT: Use u-substitution to evaluate the above integral.

Example 4:

Find the CDF of a continuous random variable with PDF

$$f_y(y) = \begin{cases} \frac{3}{8}y^2,\ 0 < y < 2 \\ 0,\ \text{elsewhere} \end{cases}$$

SOLUTION: Left as a classroom exercise.

Example 5:

Consider the function

$$f_Y(y) = \begin{cases} 0.2,\ -1 < y \le 0 \\ 0.2 + ky,\ 0 < y \le 1 \\ 0,\ \text{elsewhere} \end{cases}$$

a. Find the value of the constant k to make $f_Y(y)$ a valid PDF.

b. Find the CDF.

c. Calculate $P(-0.5 < Y < 0.3)$.

d. Calculate $P(Y < 0.3|Y < 0.5)$.

SOLUTION: Left as a classroom exercise.