Stat 121 (Mathematical Statistics I)

Lesson 3.1: Probability Distribution 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 denoted as

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

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

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

For continuous random variables, as we shall see later, 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).

In this lesson, we will define the probability density and cumulative distribution functions of continuous random variables, and articulate the dual relationship between these two functions and how these functions can be used to compute probabilities.

Definition

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

a. $f_Y(y)$ is nonnegative everywhere in the domain D, that is

$$f_Y(y) \geq 0, \ \forall y \in D$$

b. The area under the curve $f_Y(y)$in the domain is 1, that is

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

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.

Also note that for a continuous random variable, the probability that y belongs to an interval I is given by

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

In particular,

$$P(a < Y < b) = \int_a^b f_Y(y) \ dy$$

Example 3.1.1

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 $f_Y(y)$ is indeed a proability density function.

b. What is the probability that Y falls in the interval (1/2, 1)?

c. What is the probability that $Y = \frac{1}{2}$?

SOLUTION

a. First we are going check if $f_Y(y)$ is nonnegative 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 is meet. Now

$$
\begin{aligned}
\int_{y\in(0,1)} f_Y(y)\ dy &= \int_0^1 3y^2\ dy \\
&= y^3\Big|_{y=0}^{y=1} \\
&= 1^3 - 0^3 \\
&= 1
\end{aligned}
$$

which shows that the second condition is also true. Therefore, $f_Y(y)$ is a valid pdf.

b. The probability that Y falls in the interval $(\frac{1}{2}, 1)$ is given by

$$
\begin{aligned}
P(1/2 < Y < 1) &= \int_{1/2}^1 3y^2\ dy \\
&= y^3\Big|_{y=1/2}^{y=1} \\
&= 1^3 - (1/2)^3 \\
&= 7/8
\end{aligned}
$$

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

$$
\begin{aligned}
P(Y = 1/2) &= P(1/2 \leq Y \leq 1/2) \\
&= \int_{1/2}^{1/2} 3y^2\ dy \\
&= y^3\Big|_{y=1/2}^{y=1/2} \\
&= (1/2)^3 - (1/2)^3 \\
&= 0
\end{aligned}
$$

The above example shows another important property of continuous random variables. If Y is continuous, then $P(Y = y) = 0$, for all y in the domain.

Recall that the cumulative distribution function is defined for discrete random variables as $F_Y(y) = P(Y \leq y) = \sum\limits_{\forall t \leq y} P(Y = t)$.

This again means that $F_Y(y)$ accumulates all of the probability 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 nondecreasing *step* function. For continuous random variables, $F_Y(y)$ is a nondecreasing *continuous* function.

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

Remarks

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

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

b. $f_Y(y) = \frac{d}{dy}[F_Y(y)]$, provided the derivative exists.

Example 3.1.2

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 so that $f_Y(y)$ becomes a valid pdf.

b. Derive the CDF of Y.

SOLUTION

a. The second condition for a valid PDF states that

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

So, in this case,

$$\int_0^1 (y^2 + ky)dy = 1$$

Now,

$$\begin{aligned}\int_0^1 (y^2+ky)dy = 1 &\iff \left[\frac{y^3}{3}+k\frac{y^2}{2}\right]_0^1 = 1\\ &\iff \frac{1}{3}+\frac{1}{2}k = 1\\ &\iff k = \frac{4}{3}\end{aligned}$$

Thus,

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

Notice that Y assumes values between 0 and 1, thus, the function $f_Y(y) \ge 0$ which means that the first condition is also satisfied.

b. The CDF of Y is obtained as follows:

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

Next, if $0 \le y < 1$, then

$$\begin{aligned}F_Y(y) &= \int_{-\infty}^{0} 0 \ dt + \int_0^y \left(t^2+\frac{4}{3}t\right) \ dt\\ &= 0 + \left[\frac{t^3}{3}+\frac{4}{3}\frac{t^2}{2}\right]_0^y\\ &= \frac{y^3}{3}+\frac{2}{3}y^2\end{aligned}$$

Finally, if $y \ge 1$, then

$$\begin{aligned}F_Y(y) &= \int_{-\infty}^{0} 0 \ dt + \int_0^1 \left(t^2+\frac{4}{3}t\right) \ dt + \int_1^{\infty} 0 \ dt\\ &= 0 + \left[\frac{t^3}{3}+\frac{4}{3}\frac{t^2}{2}\right]_0^1 + 0\\ &= 1\end{aligned}$$

In summary,

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

Remarks

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

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

and each equals

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

Example 3.1.3

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

$$\begin{aligned} f_y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy}\left[\frac{1}{1+e^{-y}}\right] \\ &= \frac{d}{dy}(1+e^{-y})^{-1} \\ &= (-1)(1+e^{-y})^{-2}(e^{-y})(-1), \text{ by Chain Rule} \\ &= \frac{e^{-y}}{(1+e^{-y})^2} \end{aligned}$$

That is,

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

b.1. Using the CDF:

$$\begin{aligned} P(-2 < Y < 2) &= F_Y(2) - F_Y(-2) \\ &= \frac{1}{1+e^{-2}} - \frac{1}{1+e^{2}} \\ &\approx 0.881 - 0.119 \\ &= 0.762 \end{aligned}$$

b.2. Using the PDF:

$$\begin{aligned} P(-2 < Y < 2) &= \int_{-2}^{2} \frac{e^{-y}}{(1+e^{-y})^2}\, dy \\ &= -\int_{1+e^{2}}^{1+e^{-2}} u^{-2}\, du, \text{ from the substitution } u = 1 + e^{-y} \iff du = -e^{-y} \\ &= \frac{1}{u}\bigg|_{1+e^{2}}^{1+e^{-2}} \\ &= \frac{1}{1+e^{-2}} - \frac{1}{1+e^{2}} \\ &\approx 0.881 - 0.119 \\ &= 0.762 \end{aligned}$$

Example 3.1.4

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!

Learning Task

Instruction: Answer the following as indicated.

Consider the function

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

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

b. Find the CDF of Y.

c. Compute the $P(-0.5 < Y < 0.3)$.

d. Compute the $P(Y < 0.3|Y < 0.5)$