2022 SUMMER

업데이터 통계학 스터디

Chapter 3 – Continuous Random Variables 1

Continuous Random Variables

Continuous Random Variables

A continuous random variable differs from a discrete random variable in that it takes on an uncountably infinite number of possible outcomes.

Continuous Random Variables

Example

Choose a number from an interval $[0,1]$ at random.

1. What is the probability that you choose 0.5? $P(X=0.5)=0$
2. What is the probability that you choose 0.382915? $P(X=0.382915)=0$
3. What is the probability that the number is less than 0.5? $\frac{1}{2}$
4. Let X be the random number chosen from $[0,1]$. Guess what is the CDF of X?

$$F_X(x)=\begin{cases} 0 & (\infty, 0) \\ x & [0,1] \\ 1 & (1,\infty) \end{cases}$$

5. Draw the CDF of X.

Continuous Random Variables

The PMF of discrete random variable X is $P(X = x)$.

However, $P(X = x) = 0$ for continuous random variables. i.e., the probability that X takes on any value x equals 0.

This time: The probability distribution of continuous random variable X is defined via $P(a < X \leq b)$ and it means the probability that X falls in some interval $(a, b]$.

Continuous Random Variables

Definition

A random variable X is called a continuous random variable if its CDF $F_X(x)$ is a continuous function for all $x \in \mathbb{R}$.

NOTE : In discrete case, the discontinuity points have positive probability. However, the probability mass is zero for each point if the random variable is of continuous type. Let us define probability function for continuous random variable that works as probability mass function of discrete random variable.

Probability Density Function

We want to make the function $f_X(x)$ such that the probability of an interval is the area under the curve. In other words, for any interval (a, b], the pdf of X, $f_X(x)$ satisfies

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx$$

Probability Density Function

For each $x \in S_X$, consider the probability of a short interval

$$P(x < X \leq x + \Delta)$$

The probability can be computed by

$$\Delta(width) * f_X(x)(height)$$

approximately. Thus, the probability density function is defined as

$$f_X(x) = \lim_{\Delta \to 0} \frac{P(x < X \leq x + \Delta)}{\Delta}$$

Probability Density Function

$$f_X(x) = \lim_{\Delta \to 0} \frac{P(x < X \leq x + \Delta)}{\Delta}$$

$$= \lim_{\Delta \to 0} \frac{F_X(x+\Delta) - F_X(x)}{\Delta}$$

$$= F'_X(x)$$

$$\therefore f_X(x) = F'_X(x)$$

Probability Density Function

Definition.

Let X be a continuous random variable. Then the probability density function (PDF) of X is a function $f_X(x)$ defined on real numbers such that

$F_X(x)$ is non-decreasing

1. $f_X(x)$ is non-negative for any $x \in \mathbb{R}$.

2. The area under the curve $f_X(x)$ is 1, that is:

$$\int_{\mathbb{R}} f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

Probability Density Function

We call $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ the support of $f_X(x)$. Which means $f_X(x) = 0$ for any x outside S_X. Thus, we can write

$$\int_{\mathbb{R}} f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_{S_X} f_X(x)\,dx = 1$$

3. For any $\mathrm{A} \subseteq S_X$,

$$P(X \in A) = \int_A f_X(x)\,dx$$

Example 3.1

Let X be a continuous random variable whose PDF is:

$$f_X(x) = 3x^2, \qquad (0 < x < 1)$$

1. Verify that $f_X(x)$ is a valid PDF. ① non-negative ✓ ② $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ ✓ $\because \int_0^1 3x^2\,dx = 1$

2. What is the probability that X falls between $1/2$ *and* 1 ?

   That is, what is $P(1/2 < X < 1)$? $\int_{\frac{1}{2}}^{1} 3x^2\,dx = \frac{7}{8}$

3. What is $P(X = 1/2)$? 0

NOTE: $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$.

Example 3.2

Let X be a continuous random variable whose PDF is:

$$f_X(x) = \frac{x^3}{4}, \qquad (0 < x < c)$$

What value of the constant c makes $f_X(x)$ a valid PDF?

$$\int_0^c \frac{x^3}{4}\,dx = \left[\frac{x^4}{16}\right]_0^c = \frac{c^4}{16} = 1, \qquad \therefore c = 2$$

Cumulative Distribution Function

CDF of a random variable X always exists. ; $P(X \leq x)$.

Let X is a continuous random variable (c.r.v.). Then,

$$F_X(x) = P(X \leq x) = P(-\infty < X < x) = \int_{-\infty}^{x} f_X(t)\,dt$$

For $x \in \mathbb{R}$.

NOTE: For c.r.v., $F_X(x)$ is a non-decreasing continuous function.

Example 3.3

Let X be a continuous random variable whose PDF is:

$$f_X(x) = 3x^2, \qquad (0 < x < 1)$$

What is the cumulative distribution function $F_X(x)$?

$$F_X(x) = \begin{cases} x^3 & (0, 1) \\ 0 & (-\infty, 0] \\ 1 & [1, \infty) \end{cases} = x^3 \cdot \underbrace{I_{(0,1)}(x)}_{\text{Indicator}} + \underbrace{I_{(1,\infty)}(x)}_{\text{function}}$$

Indicator function

지시 함수

: 특정 집합에 특정 값이 포함되는지를 지시

def

$$\underset{\text{집합}}{I_A}(\underset{\text{값}}{x}) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Example 3.4

Let X be a continuous random variable whose PDF is:

$$f_X(x) = \frac{x^3}{4}, \qquad (0 < x < 2)$$

What is the cumulative distribution function $F_X(x)$?

$$F_X(x) = \begin{cases} \frac{1}{16}x^4 & (0, 2) \\ 0 & (-\infty, 0] \\ 1 & [2, \infty) \end{cases} \quad = \frac{1}{16}x^4 \cdot I_{(0,2)}(x) + I_{(2,\infty)}(x)$$

Mean (Expectation value)

Definition

Let X be a continuous random variable with its PDF $f_X(x)$. The expected value or mean of a continuous random variable X is:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

Let $u(x)$ is a real valued function. Then the expected value of $u(X)$ is:

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

Variance / Standard Deviation

<u>Definition</u>

The variance of a continuous random variable X is:

$$\sigma_X{}^2 = Var(X) = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \, f_X(x) \, dx$$

The **standard deviation** of a continuous random variable X is:

$$\sigma_X = \sqrt{Var(X)}$$

Variance / Standard Deviation

We can use the shortcut formula for variance.

$$E[(X-\mu_X)^2] = E(X^2) - {\mu_X}^2$$

And,

$$Var(aX+b) = a^2 Var(X)$$

Variance

Linearity of Expectation Value

Theorem 선형성

Suppose that X is a random variable with finite expectation. Then the following statements are true.

1. If c is a constant in the interval of X, then $E(c) = c$.
2. If c is a constant and u is a function, then

$$E[cu(X)] = cE[u(X)].$$

ex. $E(3X^2 + 6X + 1)$

$= 3E(X^2) + 6E(X) + 1$

3. For any positive integer k,

$$E\left[\sum_{i=1}^{k} c_i u_i(X)\right] = \sum_{i=1}^{k} c_i E[u_i(X)]$$

Example 3.5

Suppose X is a continuous random variable with the following probability density function:

$$f_X(x) = 3x^2 \quad (0 < x < 1)$$

1. What is the mean of X ? $E(X) = \int_0^1 x \cdot 3x^2 \, dx = \left[\frac{3}{4} x^4 \right]_0^1 = \frac{3}{4}$
2. What is the variance of X ?

$$E(X^2) = \int_0^1 x^2 \cdot 3x^2 \, dx = \left[\frac{3}{5} x^5 \right]_0^1 = \frac{3}{5}$$

$$\therefore \ Var(X) = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$$

Example 3.6

Let Y be a continuous random variable with PDF

$$g_Y(y) = 2y \quad (0 < y < 1)$$

1. What is the mean of Y ?
2. What is the variance of Y ?

$E(Y) = \int_0^1 y \cdot 2y \, dy = \left[\frac{2}{3}y^3\right]_0^1 = \frac{2}{3}$

$E(Y^2) = \int_0^1 y^2 \cdot 2y \, dy = \left[\frac{2}{4}y^4\right]_0^1 = \frac{1}{2}$

$Var(Y) = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$

Moment Generating Function (MGF)

적률

* $E((X-\mu)^k)$: k-th central moment
중심 적률

Definition

The moment generating function of a continuous random variable X, if it exists, is:

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx\,, \qquad t \in (-h, h), \qquad \exists_h > 0$$

Differentiating the moment generating function provides us with a way of finding the mean and the variance:

$$\mu_X = E(X) = M_X{}'(0), \qquad \sigma_X{}^2 = Var(X) = M_X{}''(0) - \{M_X{}'(0)^2\}$$

Example 3.7

$* \text{ Gamma}(\alpha, \beta)$

$\sim M_X(t) = \frac{1}{(1-\beta t)^{\alpha}}$

Suppose X is a continuous random variable with the following PDF:

$$f_X(x) = xe^{-x} \quad (0 \leq x < \infty)$$

$\sim \text{Gamma}(2,1) \rightarrow M_X(t) = \frac{1}{(1-t)^2}, \ \mu_X = 2$

Use the MGF, find the mean and variance of X.

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \cdot xe^{-x}\,dx$$

$$= \int_0^{\infty} x \cdot e^{x(t-1)}\,dx \leftarrow \text{Improper Integral } t-1<0$$

$$= \left[\frac{1}{t-1} xe^{x(t-1)} - \frac{1}{(t-1)^2} e^{x(t-1)}\right]_0^{\infty}$$

$$= \frac{1}{(1-t)^2}$$

$M_X'(t) = \frac{2}{(1-t)}$

$\mu_X = M_X'(0) = 2$

Example 3.8

* exponential distribution

$X \sim Exp(\lambda)$, $f_X(x) = \lambda e^{-\lambda x}$, $M_X(t) = \dfrac{\lambda}{\lambda - t}$

평행이동

$E(X) = \dfrac{1}{\lambda}$, $Var(X) = \dfrac{1}{\lambda^2}$

Suppose X has the PDF:

$P(X \le q_1) = .25$
$P(X \le q_2) = .5$
$P(X \le q_3) = .75$

$$f_X(x) = e^{-x-1} \quad (-1 < x < \infty)$$

1. P(X>1) $= \int_1^\infty e^{-x-1}\,dx = \left[-e^{-x-1}\right]_1^\infty = 1$
2. $M_X(t)$ $= E(e^{tX}) = \int_{-1}^\infty e^{tx} \cdot e^{-x-1}\,dx = \int_{-1}^\infty e^{(t-1)x-1}\,dx = \left[\frac{1}{t-1} e^{(t-1)x-1}\right]_{-1}^\infty = \dfrac{e^{-t}}{1-t}$
3. μ_X $= M_X'(0) = \left.\dfrac{-e^{-t}(1-t) + e^{-t}}{(1-t)^2}\right|_{t=0} = \left.\dfrac{te^{-t}}{(1-t)^2}\right|_{t=0} = 0$, $E(X^2) = M_X''(0) = \left.\dfrac{(-te^{-t} + e^{-t})(1-t)^2 + te^{-t} \cdot 2(1-t)}{(1-t)^4}\right|_{t=0} = 1$
4. σ_X $= E(X^2) - \{E(X)\}^2 = 1$
5. $F_X(x)$ $= \int_{-1}^x f_X(t)\,dt = \int_{-1}^x e^{-t-1}\,dt = \left[-e^{-t-1}\right]_{-1}^x = -e^{-x-1} + 1$
6. q_1, q_2, q_3