Handout[1] 5

Oct 2, 2022

[Topics]:
Continuous Random Variables

1 Continuous Random Variable

Consider the lifetime of a light bulb. It is a random variable that takes all the nonnegative values, i.e, [0, $+\infty$). Such random variable is not discrete, we call it "continuous".

Think adout another exmaple, there is a ruler from 0cm to 10cm. You pick a point on the ruler, what is probability that you pick exactly at 5cm? It is 0 since a point shares 0 proportion on an interval. However, if now we are considering the probability of picking a point that lies in the interval from 5cm to 6cm, then we can use the proportion of length of the desired interval over the whole length to represent the probability. Now we get $\frac{1cm}{10cm} = \frac{1}{10}$.

Basicly, for a continuous random varibale X, the exact probability for $X = x$ is 0. In order to measure the probability, we might assign a number for each x, say $f(x)$, it is not the probability $P(X = x)$. However, if we integral it on an interval, we want the integral to represent the probability of the value is on the interval.

Definition 1.1 Continuous Random Variable : A random variable X is a continuous random variable if it takes uncountable number of possible values.

Usually, continuous random variable X takes values on an interval. For example, the weight of an apple may take the possible values on [100, 300] grams. The lifetime of a light bulb may take values on [0,$+\infty$).
If you refer to the definition 3.2.1 on page 101 of the textbook(DeGroot, 4th edition), you may realize that it is a different way to define. Recall that for an integral, a countable number of points do not change the integral value. For example, for a Bernoulli random variable, we can assign $f(1) = p$, $f(0) = 1 - p$, and $f(x) = 0$ for all $x \neq 0, 1$. Then f is nonnegative, defined on the whole real line $\mathbb{R}$. However and the probability that X takes a value in the interval is not the integral of f over the interval. $P(X \in [-1, 2])$ should be 1, but $\int_{-1}^{2} f(x)dx = 0$. It does not satisfy the definiton 3.2.1 on textbook for continuous random variable, hence Bernoulli random variable is discrete.

Example 1.1 Uniform Distribution We say X is uniformly distributed if it takes equal probabilities on each interval with the same length. If $X \sim$Uniform[0, 1], it means the probability that X is on [0, 0.2] and the probability that X is on [0.6, 0.8] are the same.

2 Cumulative Distribution Functions(cdf) for Continuous Random Variables

Definition 2.1 cdf : Let X be a continuous random variable, then the cumulative distribution function(cdf) F is defined by:

$$F(x) = P(X \leq x).$$

It is the same as cdf for discrete random variables. The properties also hold, and there is a further property (5) for continuous random variables:

$$\begin{aligned}
&(1) \quad 0 \leq F(x) \leq 1, \\
&(2) \quad F(x) \text{is nondecreasing, that is, if } x_1 < x_2, \text{ then } F(x_1) \leq F(x_2), \\
&(3) \quad lim_{x \to -\infty} F(x) = 0, \\
&(4) \quad lim_{x \to +\infty} F(x) = 1. \\
&(5) \quad F(x) \text{ is a continuous function of x, and F(x) has a derivative function } f(x).
\end{aligned}$$

[1]This handout is made by Hongkai Wang for Econ 329 Economic Statistics. It is adapted from Professor Wiseman's lectures, however, all errors are mine.

Actually, property (5) is the essential property for continuous random varaibles. Recall for discrete random variables, the cdf $F(x)$ is a step function.

The $f(x)$ in property (5) is known as the **density function** of the random variable X.

3 Probability Density Function(pdf) for Continuous Random Variables

Definition 3.1 (pdf) : Let $F(x)$ be the cdf of a continuous random variable X. Then $f(x) = F'(x)$ is called a probability density function (pdf) of the random variable X.

We call $f(x)$ "a pdf" rather than "the pdf" because the pdf is not unique. As we can change a countable number of values of points $f(x)$, while the integral remians unchanged.

Recall in the beginning, for a continuous random variable X, we want to assign a number $f(x)$ to each x, such that the integral $\int_a^b f(x)dx$ represents the probability that X takes the value on the interval [a, b]. So what relations $f(x)$ and $F(x)$ should satisfy?

Relations Between cdf and pdf :

$$(1) \quad F'(x) = f(x),$$
$$(2) \quad F(x) = \int_{-\infty}^{x} f(x)dx.$$

Based on the definition of $F(x)$, we know $P(x_1 \leq X \leq x_2)$=$F(x_2) - F(x_1)$=$\int_{x_1}^{x_2} f(x)dx$. This is consistent with what we desired.

Let's take a review on what we have done. For a continuous random variable X, we are unable to define the probability at each x. But the cdf $F(x)$ is well defined. We want to assign each x a number $f(x)$, such that it can be used to measure the probability of a continuous random variable. Specifically, we want $\int_a^b f(x)dx$ to represent the probability $P(X \in [a, b])$. Some how we find $f(x) = F'(x)$, and we verified that for a<b,

$$\begin{aligned}
\int_a^b f(x)dx &= \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx \\
&= F(b) - F(a) \\
&= P(X \leq b) - P(X \leq a) \\
&= P(X \in [a, b])
\end{aligned}$$

It seems $f(x) = F'(x)$ is good. But we need to verify one more thing: whether $f(x_1) > f(x_2)$ represents it is more likely that X taking the value x_1 to happen than X taking the value x_2. Though we said that the exact probability of X takes a value x is 0, it does not mean we are unable to compare the two probabilities. Think about the height of the students in the Econ 329 class. The height is a continuous random variable that takes values on, say, [0, 10] feets. It is reasonable to say $X = 6$ is more likely to happen than $X = 8$.

For $f(x) = F'(x)$, consider a small interval $(x - \epsilon, x + \epsilon)$:

$$\begin{aligned}
f(x) &= F'(x) \\
&= lim_{\epsilon \to 0} \frac{F(x+\epsilon) - F(x-\epsilon)}{2\epsilon} \\
&= lim_{\epsilon \to 0} \frac{1}{2\epsilon} P(X \in [x-\epsilon, x+\epsilon])
\end{aligned}$$

If we fix ϵ, then $f(x_1) > f(x_2)$ reveals $P(X \in [x_1 - \epsilon, x_1 + \epsilon]) > P(X \in [x_2 - \epsilon, x_2 + \epsilon])$. For ϵ small enough, it can be considered as of $X = x_1$ is more likely to happen than $X = x_2$.

Now everything is good. We find a f to represent the "probability density" of a continuous random variable.

Property 3.1 (Properties of pdf) : Let $f(x)$ be a pdf of a continuous random variable, then :

$$(1) \quad f(x) \geq 0,$$
$$(2) \quad \int_{-\infty}^{+\infty} f(x)dx = 1.$$

Proof :
(1) $f(x) \geq 0$ because it is the derivative of a nondecreasing function $F(x)$.
(2) $\int_{-\infty}^{+\infty} f(x)dx = \lim_{x\to+\infty} F(x) = 1$.

Example 3.1 Suppose the cdf of a certain continuous random variable is given by:

$$F(x) = \begin{cases} 0 \quad if \quad x < 0 \\ x^2 \quad if \quad 0 \leq x < 1 \\ 1 \quad if \quad x \geq 1 \end{cases}$$

What is a pdf of X?

$$f(x) = F'(x) = \begin{cases} 0 \quad if \quad x < 0 \\ 2x \quad if \quad 0 \leq x < 1 \\ 0 \quad if \quad x \geq 1 \end{cases}$$

In fact, we can assign any value to $x = 0$ and $x = 1$. If we assign 0 on both, then we get

$$f_1(x) = \begin{cases} 2x \quad if \quad 0 \leq x < 1 \\ 0 \quad otherwise \end{cases}$$

In fact, even for $x \in [0, 1)$, you can assign a countable number of them to have differnet values as $2x$. For example, let

$$f_2(x) = \begin{cases} 2x \quad if \quad x \in [0, 0.5) \cup (0.5, 1) \\ 100 \quad if \quad x = 0.5 \\ 0 \quad otherwise \end{cases}$$

$f_2(x)$ is also a valid pdf for X.

4 Basic Formula

4.1 For any random variable (whether discrete or continuous) X, let $F(x)$ be the cdf of X. Then for any $a < b$,

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

Proof : $\{X \leq b\}$=$\{X \leq a\} \cup \{a < X \leq b\}$. And $\{X \leq a\}$, $\{a < X \leq b\}$ are disjoint. Then

$$\begin{aligned} P(X \leq b) &= P(X \leq a) + P(a < X \leq b) \\ \Rightarrow P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a). \end{aligned}$$

4.2 For continuous random variable X, let $F(x)$ be the cdf and $f(x)$ be a pdf. Then for $a < b$,

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

Proof : By relation (2) :

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

Then

$$F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx.$$

4.3 For continuous random variable X, for any real value b,

$$P(X = b) = 0.$$

Proof :

$$\begin{aligned} P(X = b) &= \lim_{n \to 0} P(b - \frac{1}{n} < X \leq b) \\ &= \lim_{n \to 0} \int_{b-\frac{1}{n}}^{b} f(x)dx \\ &= \int_{b}^{b} f(x)dx \\ &= 0. \end{aligned}$$

Note that this is only true for continuous random varibales.

Wrap up : Suppose X is a continuous random varibale with probability density function $f(x)$. Then for any $-\infty < a < b < +\infty$,

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_{a}^{b} f(x)dx.$$

i.e, the probabilities of the intervals (a, b], [a, b], [a, b), (a, b) are the same.

4.4 For a continuous random varibale X, let $F(x)$ be the cdf and $f(x)$ be a pdf. Then for any a,

$$\begin{aligned} P(X > a) = P(X \geq a) &= 1 - F(a) \\ &= 1 - \int_{-\infty}^{a} f(x)dx \\ &= \int_{a}^{+\infty} f(x)dx. \end{aligned}$$

Also,

$$P(X < a) = P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)dx.$$

5 Geometric Interpretation

For a continuous random varibale X with pdf $f(x)$, $P(a < X < b)$ is just the area of the shaded one:

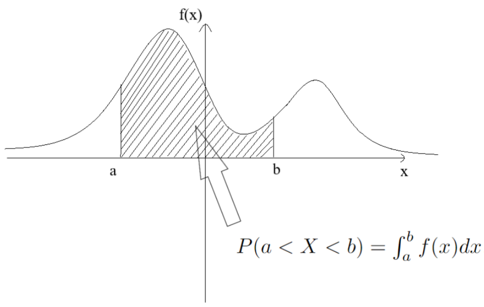

Figure 1: pdf of continuous randmom varibale X

Now think about what does $\int_{-\infty}^{+\infty} f(x)dx = 1$ mean? It means on the pdf graph, the area under the pdf curve should be 1.

Example 5.1 Consider the uniform random variable $X \sim$Uniform$[a, b]$. What is its pdf?
For the uniform random variable, the pdf $f(x)$ is identical for all x on $[a, b]$. The length of the interval is $b - a$, hence in order to make the area under the curve to be 1, we need $f(x)$ to be $\frac{1}{b-a}$ for $x \in [a, b]$.

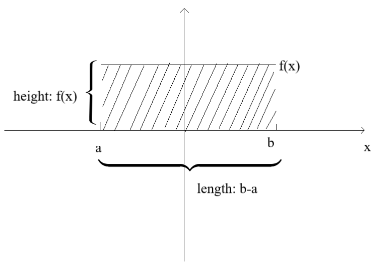

Figure 2: pdf of $X \sim Uniform[a, b]$

6 Quantile

Definition 6.1 (Quantile) : For a random variable X with cdf $F(x)$, the p quantile of X, denoted as $F^{-1}(p)$, is defined to be the smallest value x such that $F(x) \geq p$, where $0 < p < 1$.

Quantile of continuous random varibale For continuous random variable X, its cdf $F(x)$ is continuous and one-to-one maps from x to $F(x)$. Hence the p quantile of X is the inverse function $F^{-1}(p)$.

Example 6.1 Let $X \sim Uniform[a, b]$. What are the 0.25, 0.5, 0.75 quantiles of X?
Solution : $F(x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$. To figure out the inverse function of $F(x)$, first set $F(x) = \frac{x-a}{b-a} = y$, then $x = a + y(b - a) = yb + (1 - y)a$. i.e, $F^{-1}(y) = yb + (1 - y)a$.
Now $F^{-1}(0.25) = 0.25b + 0.75a$, $F^{-1}(0.5) = 0.5b + 0.5a$, and $F^{-1}(0.75) = 0.75b + 0.25a$.

Quantile of discrete random varibale For discrete random variable X, we need to refer to the cdf table to determine the quantile. We should find the smallest x such that $F(x) \geq p$.

Example 6.2 Let X be a discrete random variable with pdf and cdf as shown in the table. What are the 0.25, 0.5, 0.75, 0.95 quantiles of X?

x	0	1	2	3	4	5
f(x)	0.1681	0.3602	0.3087	0.1323	0.0284	0.0024
F(x)	0.1681	0.5283	0.8370	0.9693	0.9977	1

Solution : The 0.25 quantile is $x = 1$, the 0.5 quantile is $x = 1$, the 0.75 quantile is $x = 2$, the 0.95 quantile is $x = 3$.

7 Examples of Continuous Random Variables

7.1 Uniform Random Variable

Let X be a uniform random variable on [0, 1], denote as $X \sim Uniform[0, 1]$.

The pdf is given by:

$$f(x) = \begin{cases} 1 & if \quad 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

The cdf is given by:

$$F(x) = \begin{cases} 0 & if \quad x < 0 \\ x & if \quad 0 \leq x \leq 1 \\ 1 & if \quad x > 1 \end{cases}$$

7.2 Exponential Distribution

Let X be exponentially distributed with parameter $\lambda(\lambda > 0)$, denote as $X \sim Exponential(\lambda)$.

The pdf is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & if \quad x \geq 0 \\ 0 & otherwise \end{cases}$$

The cdf is given by:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & if \quad x \geq 0 \\ 0 & otherwise \end{cases}$$

Recall $F(x) = \int_{-\infty}^{x} f(x)dx$. Then if $x < 0$,

$$\int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} 0dx = 0.$$

If $x \geq 0$, then

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(x)dx \\ &= \int_{-\infty}^{0} 0dx + \int_{0}^{x} \lambda e^{-\lambda x}dx \\ &= \int_{0}^{x} \lambda e^{-\lambda x}dx \\ &= -e^{-\lambda x}\Big|_{0}^{x} \\ &= 1 - e^{-\lambda x} \end{aligned}$$

8 Exercises

8.1 A continuous random variable X, with given pdf:

$$f(x) = \begin{cases} c(3-x) & if \quad 0 < x < 3 \\ 0 & otherwise \end{cases}$$

(1) Find c.
(2) Find cdf.

Solution :
(1) $\int_0^3 c(3-x)dx = (3cx - \frac{c}{2}x^2)\Big|_0^3 = \frac{9}{2}c = 1. \Rightarrow c = \frac{2}{9}$.
(2) If $x \leq 0$, then F(x)=0.
If $0 < x < 3$, then $F(x) = \int_{-\infty}^{x} f(s)ds = \int_0^x \frac{2}{9}(3-s)ds = \frac{2}{9}\ (3s - \frac{s^2}{2})\Big|_0^x = \frac{2}{3}x - \frac{1}{9}x^2$.
If $x \geq 3$, then $F(x) = 1$.

8.2 Consider the random variable X with cdf given as:

$$F(x) = \begin{cases} 0 \quad if \quad x < 0 \\ 1 - e^{-2x} - 2xe^{-2x} \quad if \quad x \geq 0 \end{cases}$$

Find a pdf of X.

Solution : If $x < 0$, then $F'(x) = 0$.
If $x \geq 0$, then $F'(x) = 2e^{-2x} + (-2e^{-2x} + 4xe^{-2x}) = 4xe^{-2x}$.
Hence

$$f(x) = \begin{cases} 0 \quad if \quad x < 0 \\ 4xe^{-2x} \quad if \quad x \geq 0 \end{cases}$$

Good luck on Your Midterm!