Math 170E

Lecture Notes Section 3.1 [*†]

Continuous random variables

Instructor: Swee Hong Chan

NOTE: Materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

[*]Version date: Wednesday 27^{th} January, 2021, 12:53.

[†]This notes is based on March Boedihardjo's and Jamie Haddock's notes from the past quarters, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

1 Continuous random variable: Definition

A random variable X is **continuous** if its cumulative distribution function is continuous. (Recall that the cumulative distribution function is $F : \mathbb{R} \to [0, 1]$, where

$$F(x) \;:=\; P(X \leq x) \qquad \text{for } \; x \in (-\infty, \infty)\,.)$$

2 Continuous random variable: Example

Let X be a real number that you select at random from from the interval $[0, 1]$ (i.e., a real number that is greater than 0 and less than 1). Then

- The probability that the random number X is less than -5 is equal to 0.
- The probability that the random number X is less than 10 is equal to 1.
- The probability that the random number X is less than $\frac{1}{2}$ is equal to $\frac{1}{2}$.

In general, for any real number x, the cdf of X is

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

This is a continuous function (see picture (BT)). Thus X is a **continuous random variable**.

3 Probability density function: Definition

The **probability density function** of a continuous random variable X is the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) \quad = \quad \frac{d}{dx} F(x).$$

where $F(x)$ is the cumulative distribution function of X.

In theory, the function $F(x)$ does not have to be differentiable.

However, in this course we will only deal with nice functions, so $F(X)$ will be **differentiable except for finitely many points**.

4 Probability density function: Example

Let X be the real number selected uniformly at random from $[0,1]$. The cdf of X is

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

The pdf of X is then equal to

$$f(x) \quad = \quad \frac{d}{dx} F(x) \quad = \quad \begin{cases} 1 & \text{for} \quad 0 < x < 1; \\ 0 & \text{otherwise.} \end{cases}$$

Note that the derivative is not defined at 0 and 1. We will without loss of generality assume that $f(0) = f(1) = 0$.

5 Probability density function: Properties

- $f(x) \geq 0$ for any real number x.
- $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$
- For any real numbers a and b,

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

- The cdf of the random variable X is then

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

for any real number x.

6 Probability density function vs probability mass function

Probability density function is defined for **continuous** RVs, and is equal to

$$f(x) \quad = \quad \frac{d}{dx}F(x).$$

Probability mass function is defined for **discrete** RVs, and is equal to

$$f(x) \quad = \quad P(X = x).$$

Important: For continuous RVs, $P(X = x)$ **is always equal to 0**, and the probability density function $f(x)$ **can be greater than 1**.

7 Uniform random variable (continuous)

Let a, b be real numbers with $a < b$.

The **uniform random variable** X on $[a, b]$ is the random number chosen uniformly from the interval $[a, b]$.

This random variable is sometimes denoted $U(a, b)$, and is called **rectangular RVs**.

8 Uniform random variable: pdf and cdf

This random variable has probability density function

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for} \quad a < x < b; \\ 0 & \text{otherwise,} \end{cases}$$

and cumulative distributive function

$$F(x) = \begin{cases} 0 & \text{if } x < a; \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b; \\ 1 & \text{if } x > b. \end{cases}$$

The support of this random variable is $[a, b]$.

9 Continuous RVs: Another example

Let X be a continuous random variable with pdf $f(x) = 4x$ for $0 < x < \sqrt{\frac{1}{2}}$. Compute the probability that X is strictly between $\frac{1}{4}$ and $\frac{1}{2}$.

Answer: The cdf of X is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0; \\ \int_0^x 4y\,dy = 2x^2 & \text{for } 0 \le x \le \sqrt{\frac{1}{2}}; \\ 1 & \text{for } x > \sqrt{\frac{1}{2}}. \end{cases}$$

Thus the answer is

$$\begin{aligned} P(1/4 < X < 1/2) &= F(1/2) - F(1/4) \\ = 2\left(\frac{1}{2}\right)^2 - 2\left(\frac{1}{4}\right)^2 &= \frac{1}{2} - \frac{1}{8} = \frac{3}{8}. \end{aligned}$$

10 Continuous RVs: Mean, variance, moment generating function

Let X be a continuous random variable with pdf $f(x)$.

The mean of X is

$$\mu \quad = \quad E[X] \quad = \quad \int_{-\infty}^{\infty} x f(x)\,dx.$$

The variance of X is

$$\sigma^2 \quad = \quad \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \quad = \quad \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2.$$

The standard deviation of X is

$$\sigma \quad = \quad \sqrt{\sigma^2}.$$

The moment generating function of X is

$$M(t) \quad = \quad \int_{-\infty}^{\infty} e^{tx} f(x)\, dx.$$

Most properties of expectations and moment generating functions from before will still hold for continuous random variables. This includes

- Linearity of expectations;
- $M(0) = 1$;
- $E[X^r] = M^{(r)}(0)$ for all $r \geq 1$.

11 Uniform RVs: mean, variance mgf

Theorem 1. *Let X be the uniform random variable on the interval $[a, b]$. The moment generating function for X is*

$$M(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & \textit{for} \;\; t \neq 0; \\ 1 & \textit{for} \;\; t = 0. \end{cases}$$

The mean and variance of X is given by

$$\mu = \frac{a+b}{2}; \qquad \sigma^2 = \frac{(b-a)^2}{12}.$$

12 Uniform RVs: Proof

The mean of X is equal to (BT)

$$\begin{aligned} \mu &= \int_a^b x \frac{1}{b-a} \, dx &= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\ &= \frac{1}{2(b-a)} (b^2 - a^2) &= \frac{b+a}{2}. \end{aligned}$$

The second moment of X is equal to

$$\begin{aligned} E[X^2] &= \int_a^b x^2 \frac{1}{b-a} \, dx &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b \\ &= \frac{1}{3(b-a)} (b^3 - a^3) &= \frac{b^2 + ab + a^2}{3} \end{aligned}$$

The variance of X is given by

$$\begin{aligned} \sigma^2 &= E[X^2] - \mu^2 &= \left(\frac{b^2 + ab + a^2}{3} \right) - \left(\frac{b+a}{2} \right)^2 \\ &= \left(\frac{b^2 + ab + a^2}{3} \right) - \left(\frac{b^2 + 2ab + a^2}{4} \right) \\ &= \frac{b^2}{12} - \frac{2ab}{12} + \frac{a^2}{12} &= \frac{(b-a)^2}{12}. \end{aligned}$$

The moment generating function of X is, for $t = 0$, is $M(0) = 1$. For $t \neq 0$, we have

$$\begin{aligned} M(t) &= \int_a^b e^{tx} \frac{1}{b-a} \, dx = \frac{1}{b-a} \left[\frac{e^{tx}}{t} \right]_a^b \\ &= \frac{e^{tb} - e^{ta}}{t(b-a)}. \end{aligned}$$

13 Continuous RVs: Yet another example

Let X be the random variable with pdf $f(x) = |x|$ for $-1 < x < 1$.

Compute the cdf of X.

14 Continuous RVs: Yet another example, answer

For $-1 < y \le 0$, we have

$$\begin{aligned} F(y) &= P(X \le y) = \int_{-1}^{y} |x|\,dx = \int_{-1}^{y} -x\,dx \\ &= \left[-\frac{x^2}{2}\right]_{-1}^{y} = -\frac{y^2}{2} + \frac{1}{2} = \frac{1-y^2}{2}. \end{aligned}$$

For $0 < y \le 1$, we have

$$\begin{aligned} F(y) &= P(X \le y) = \int_{-1}^{y} |x|\,dx \\ &= \int_{-1}^{0} |x|\,dx + \int_{0}^{y} |x|\,dx \\ &= \int_{-1}^{0} -x\,dx + \int_{0}^{1} x\,dx \\ &= \left[-\frac{x^2}{2}\right]_{-1}^{0} + \left[\frac{x^2}{2}\right]_{0}^{y} \\ &= \frac{1}{2} + \frac{y^2}{2} = \frac{y^2+1}{2}. \end{aligned}$$

Hence the cdf of X is given by

$$F(x) \quad = \quad \begin{cases} 0 & \text{for } x \leq -1; \\ \frac{1-x^2}{2} & \text{for } -1 < x \leq 0; \\ \frac{x^2+1}{2} & \text{for } 0 < x \leq 1; \\ 1 & \text{for } x > 1. \end{cases}$$

15 Percentile: Definition

The (100p)th percentile is a number π_p such that

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

The 50th percentile is called the **median**, and is denoted by m.

The 25th, 50th, and the 75th percentiles are called the **first**, **second**, and **third quartiles**, respectively.

16 Percentile: Example

Let X be the random variable from before, with pdf $f(x) = |x|$ for $-1 < x < 1$.

Compute the first quartile, the second quartile, and the third quartile of X.

17 Percentile: Answer

Recall that we already know the cdf of this random variable from the previous example.

The second quartile for this random variable is equal to 0, since

$$F(0) \quad = \quad \frac{1-(0)^2}{2} \quad = \quad \frac{1}{2}.$$

This means that the first quartile is less than 0, and the third quartile is greater than 0.

The first quartile can now be solved by the equation

$$0.25 \quad = \quad F(\pi_{0.25}) \quad = \quad \frac{1 - \pi_{0.25}^2}{2},$$

where the last equality is because the first quartile is less than 0. Hence we have $\pi_{0.25} = -1/\sqrt{2}$.

The third quartile can now be solved by the equation

$$0.75 \quad = \quad F(\pi_{0.75}) \quad = \quad \frac{1 + \pi_{0.75}^2}{2}.$$

where the last equality is because the third quartile is greater than 0. Hence we have $\pi_{0.75} = 1/\sqrt{2}$.